

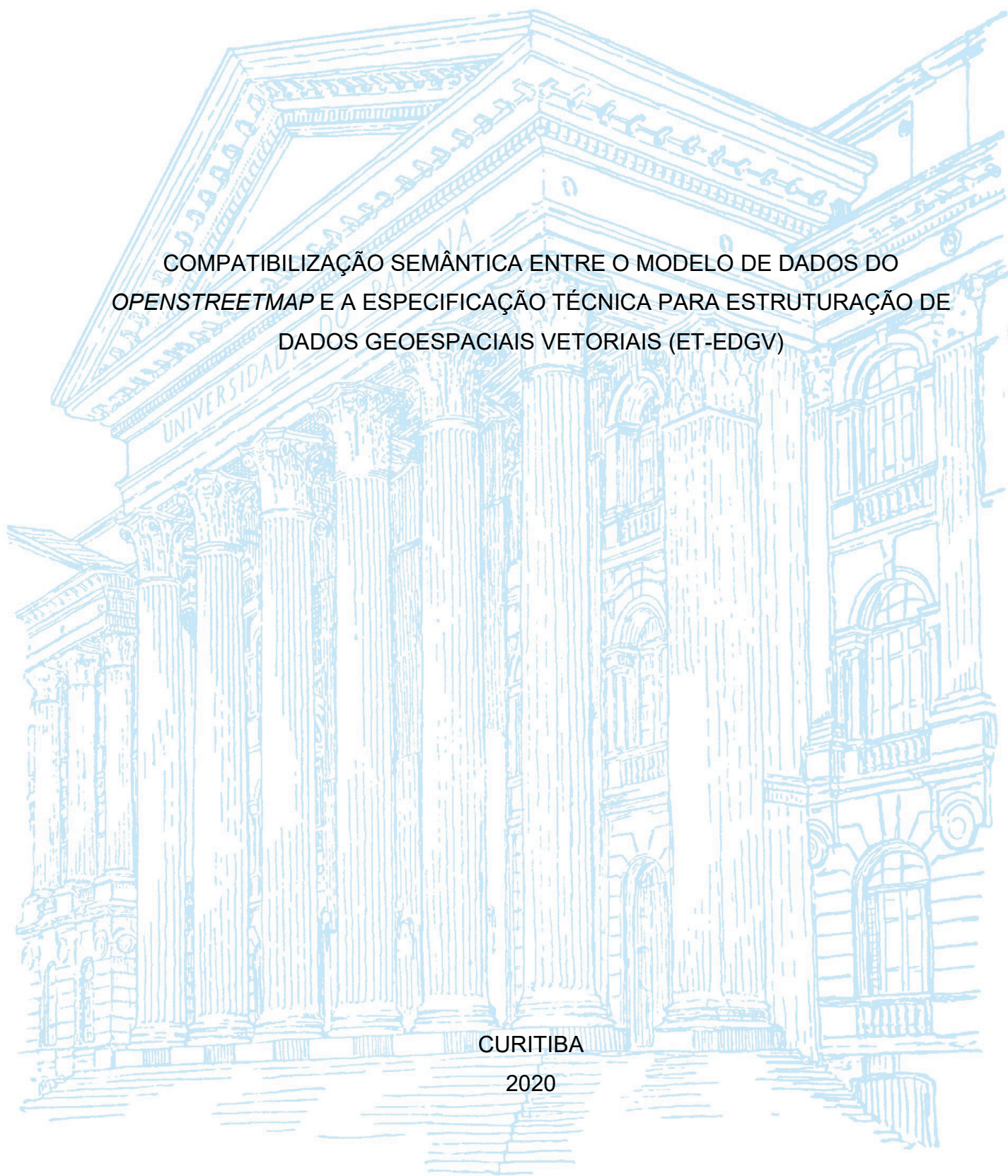
UNIVERSIDADE FEDERAL DO PARANÁ

ADRIANA ALEXANDRIA MACHADO

COMPATIBILIZAÇÃO SEMÂNTICA ENTRE O MODELO DE DADOS DO  
OPENSTREETMAP E A ESPECIFICAÇÃO TÉCNICA PARA ESTRUTURAÇÃO DE  
DADOS GEOESPACIAIS VETORIAIS (ET-EDGV)

CURITIBA

2020



ADRIANA ALEXANDRIA MACHADO

COMPATIBILIZAÇÃO SEMÂNTICA ENTRE O MODELO DE DADOS DO  
*OPENSTREETMAP* E A ESPECIFICAÇÃO TÉCNICA PARA ESTRUTURAÇÃO DE  
DADOS GEOESPACIAIS VETORIAIS (ET-EDGV)

Tese apresentada como requisito parcial  
à obtenção do título de Doutora em  
Ciências Geodésicas no curso de Pós-  
Graduação em Ciências Geodésicas,  
Setor de Ciências da Terra da  
Universidade Federal do Paraná – UFPR.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Silvana Philippi  
Camboim

CURITIBA

2020

CATALOGAÇÃO NA FONTE – SIBI/UFPR

---

M149c

Machado, Adriana Alexandria

Compatibilização semântica entre o modelo de dados do openstreetmap e a especificação técnica para estruturação de dados geoespaciais vetoriais (ET-EDGV) [recurso eletrônico]/ Adriana Alexandria Machado, 2020.

Tese (Doutorado) - Programa de Pós-Graduação em Ciências Geodésicas, Setor de Ciências da Terra da Universidade Federal do Paraná – UFPR

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Silvana Philippi Camboim

1. Carografia. 2. Semântica. 3. Processamento da informação humana. 4. Sistema de informação geográfica. I. Camboim, Silvana Philippi. II. Universidade Federal do Paraná. III. Título.

CDD526

---

Bibliotecária: Vilma Machado CRB9/1563



MINISTÉRIO DA EDUCAÇÃO  
SETOR DE CIÊNCIAS DA TERRA  
UNIVERSIDADE FEDERAL DO PARANÁ  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIAS  
GEODÉSICAS - 40001016002P6

## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIAS GEODÉSICAS da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **ADRIANA ALEXANDRIA MACHADO** intitulada: **COMPATIBILIZAÇÃO SEMÂNTICA ENTRE O MODELO DE DADOS DO OPENSTREETMAP E A ESPECIFICAÇÃO TÉCNICA PARA ESTRUTURAÇÃO DE DADOS GEOESPACIAIS VETORIAIS (ET-EDGV)**, sob orientação da Profa. Dra. SILVANA PHILIPPI CAMBOIM, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 19 de Fevereiro de 2020.

SILVANA PHILIPPI CAMBOIM

Presidente da Banca Examinadora (UNIVERSIDADE FEDERAL DO PARANÁ)

JOÃO VITOR MEZA BRAVO  
Avaliador Externo (UNIVERSIDADE FEDERAL DE UBERLÂNDIA)

LUCIENE STAMATO DELAZARI  
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

CLAUDIA ROBBI SLUTER  
Avaliador Externo (UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL)



Dedico esta tese à meu pai (*in memoriam*) pela vida, pelo incentivo e apoio para estudar, aprender e ensinar.

## **AGRADECIMENTOS**

À minha orientadora Prof<sup>a</sup>. Dr<sup>a</sup>. Silvana Philippi Camboim por guiar tão sabiamente e de modo exemplar minha trajetória acadêmica. Obrigada pelos conhecimentos, pela paciência, compreensão, incentivo e apoio constante. Não existem palavras para expressar meu respeito, admiração e gratidão.

À Prof<sup>a</sup>. Dr<sup>a</sup>. Cláudia Robbi Sluter (UFRGS), Prof. Dr. João Vitor Meza Bravo (UFU), Prof<sup>a</sup>. Dr<sup>a</sup>. Luciene Delazari (UFPR) e Prof. Dr. Tony Vinicius Moreira Sampaio (UFPR), pelas inestimáveis sugestões para melhoria da pesquisa e do manuscrito da tese.

À todos os professores do Programa de Pós Graduação em Ciências Geodésicas (PPGCG) da UFPR pelos conhecimentos compartilhados.

À todos os amigos e colegas do curso, em especial: Mônica, Rhaíssa, Cristiane, Caísse, Mário, Andrea, Eduardo, Raphael, Everton, Flávia, Gabriele, Idalécio, Vitor e Jaqueline.

À todos os funcionários do PPGCG, especialmente a Verali Mônica Kleuser e Fabiane Oshikawa, por facilitarem nossas vidas nas dificuldades do dia-a-dia.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio financeiro através da bolsa de estudos e taxa de bancada, que possibilitaram a realização da pesquisa.

Aos amigos Adélia Selpa Campestrini, Ronaldo Francisco Jorge dos Santos e Leandro Luiz dos Santos pela amizade e apoio incondicional.

Ao grande responsável por tudo isso: Deus!

“Os analfabetos do século XXI não serão aqueles que não sabem ler e escrever, mas aqueles que não sabem aprender, desaprender e reaprender.”

Alvin Toffler

## RESUMO

Em virtude do baixo investimento em cartografia no Brasil, o mapeamento topográfico oficial cobre o território de modo desigual entre as pequenas e médias escalas, além de estar muitas vezes desatualizado. Em contrapartida, o mapeamento colaborativo de plataformas como o *OpenStreetMap* (OSM) produz e disponibiliza dados geoespaciais atualizados e abertos. Desse modo, o mapeamento colaborativo se apresenta como uma alternativa relevante para atualização e completude do mapeamento topográfico oficial. Entretanto, alcançar a interoperabilidade entre dados e sistemas de geoinformação não é uma tarefa trivial devido à complexidade das informações geoespaciais e pelas várias formas através das quais ela pode ser modelada, podendo haver incompatibilidade em nível sintático e/ou semântico. Sob o aspecto semântico, é necessário garantir que os conceitos e significados dos termos de cada um dos elementos e feições sejam definidos e compartilhados adequadamente para que não haja comprometimento da qualidade das informações. Um modelo conceitual fornece uma descrição coerente e sistemática do conteúdo e organização de um conjunto de dados. Esse *schema* inclui a definição do significado das feições e demais componentes, primordiais para o intercâmbio semântico entre diferentes bancos de dados. Desse modo, a definição de um modelo conceitual, conduz a um caminho para suplantiar o desafio da interoperabilidade em nível semântico. Contudo, a falta de modelos conceituais comuns implica em problemas na troca de dados entre as organizações. Sendo assim, para que seja possível integrar o banco de dados do mapeamento colaborativo do OSM ao banco de dados do mapeamento topográfico oficial brasileiro é necessário realizar uma compatibilização semântica entre os modelos conceituais de ambos os mapeamentos. Nesse contexto, o uso de ontologias auxilia nesta questão fornecendo ferramentas para obter melhor entendimento do universo geográfico, bem como, facilitar a troca de informações entre diferentes grupos de especialistas e usuários na manipulação das entidades geográficas pelos sistemas de informações geoespaciais. As ontologias e os modelos conceituais fazem uso dos mesmos elementos conceituais e semânticos para promover a interoperabilidade entre os sistemas. Utilizando então, as ontologias de cada um dos mapeamentos, foi proposto um método híbrido de alinhamento semântico entre elas baseado em *corpus* e em conhecimento, o qual produziu uma terceira ontologia e uma base de conhecimento que permite conectar os bancos de dados do mapeamento oficial brasileiro e do OSM, bem como gerar novas ferramentas de inteligência artificial e sistemas baseados em conhecimento. Os resultados demonstraram que o alinhamento semântico é eficaz para as feições no nível básico de abstração, e que nos casos cujo nível de detalhamento e abstração é maior, é necessária a análise de contexto local e propósito de uso dos dados na classificação final dos dados. Concluiu-se que a integração dos dados do *OpenStreetMap* aos dados do mapeamento topográfico oficial brasileiro se mostrou promissora e que algumas categorias de informações podem ser mais facilmente integradas do que outras devido a questões legais e de categorização das feições.

**Palavras-chave:** integração entre bancos de dados geoespaciais. alinhamento semântico. mapeamento topográfico oficial brasileiro. *OpenStreetMap*. ET-EDGV 3.0. modelo conceitual. Ontologia. alinhamento entre ontologias. base de conhecimento. inteligência coletiva. mapeamento colaborativo. VGI.



## ABSTRACT

Due to the low investment in cartography in Brazil, the official topographic mapping covers the territory unevenly between small and medium scales, in addition to being often outdated. In contrast, collaborative mapping of platforms such as OpenStreetMap (OSM) produces and provides updated and open geospatial data. Thus, collaborative mapping presents itself as a relevant alternative for updating and completing the official topographic mapping. However, achieving interoperability between data and geoinformation systems is not a trivial task due to the complexity of geospatial information and the various ways in which it can be modeled, with incompatibility at a syntactic and / or semantic level. Under the semantic aspect, it is necessary to ensure that the concepts and meanings of the terms of each of the elements and features are properly defined and shared so that there is no compromise in the quality of the information. A conceptual model provides a coherent and systematic description of the content and organization of a dataset. This schema includes the definition of the meaning of features and other components, which are essential for the semantic exchange between different databases. In this way, the definition of a conceptual model, leads to a way to overcome the challenge of interoperability at the semantic level. However, the lack of common conceptual models implies problems in the exchange of data between organizations. Therefore, in order to be able to integrate the OSM collaborative mapping database to the official Brazilian topographic mapping database, it is necessary to perform a semantic compatibility between the conceptual models of both mappings. In this context, the use of ontologies assists in this issue by providing tools to obtain a better understanding of the geographic universe, as well as, facilitating the exchange of information between different groups of specialists and users in the manipulation of geographic entities by geospatial information systems. Ontologies and conceptual models make use of the same conceptual and semantic elements to promote interoperability between systems. Using then, the ontologies of each of the mappings, a hybrid method of semantic alignment between them, based on corpus and knowledge was proposed, which produced a third ontology and a knowledge base that allows to connect the databases of the official Brazilian mapping and OSM, as well as generating new artificial intelligence tools and knowledge-based systems. The results showed that semantic alignment is effective for features at the basic level of abstraction, and that in cases where the level of detail and abstraction is higher, it is necessary to analyze the local context and purpose of using the data in the final classification of the data. It was concluded that the integration of OpenStreetMap data with data from the official Brazilian topographic mapping proved to be promising and that some categories of information can be more easily integrated than others due to legal and categorization issues.

**Keywords:** integration between geospatial datasets. semantic alignment. Brazilian official topographic mapping. OpenStreetMap. ET-EDGV 3.0. conceptual models. Ontologies. alignment between ontologies. knowledge base. collective intelligence. collaborative mapping. volunteered geographic information (VGI).

## LISTA DE FIGURAS

FIGURA 1 – RELACIONAMENTO ENTRE NÍVEL DE ABSTRAÇÃO, OBJETO & CATEGORIA E TIPO DE RACIOCÍNIO .....	28
FIGURA 2 – TIPOS DE ONTOLOGIAS .....	35
FIGURA 3 – COMBINAÇÃO DE ONTOLOGIAS.....	38
FIGURA 4 – ALINHAMENTO DE ONTOLOGIAS .....	39
FIGURA 5 – MAPEAMENTO DE ONTOLOGIAS.....	40
FIGURA 6 – INTEGRAÇÃO DE ONTOLOGIAS.....	41
FIGURA 7 – NÍVEIS DE ABSTRAÇÃO DE APLICAÇÕES GEOESPACIAIS .....	48
FIGURA 8 – NOTAÇÃO GRÁFICA PARA AS CLASSES DO MODELO OMT-G.....	52
FIGURA 9 – NOTAÇÃO GRÁFICA PARA GEO-CAMPOS.....	52
FIGURA 10 – NOTAÇÃO GRÁFICA PARA GEO-OBJETOS.....	53
FIGURA 11 – RELACIONAMENTOS.....	54
FIGURA 12 – GENERALIZAÇÃO ESPACIAL.....	56
FIGURA 13 – AGREGAÇÃO.....	57
FIGURA 14 – GENERALIZAÇÃO CARTOGRÁFICA NO MODELO OMT-G .....	58
FIGURA 15 – DIAGRAMA DE CLASSES NA ET-EDGV 3.0.....	68
FIGURA 16 – TABELA DA CLASSE COMPLEXO AEROPORTUÁRIO .....	69
FIGURA 17 – NÓS NO OSM.....	86
FIGURA 18 – CAMINHO ABERTO NO OSM.....	87
FIGURA 19 – CAMINHO FECHADO E ÁREA NO OSM.....	87
FIGURA 20 – MEMBRO, PAPEL E RELAÇÃO NO OSM .....	88
FIGURA 21 – ETIQUETAS NO OSM.....	88
FIGURA 22 – DESCRIÇÕES SEMÂNTICAS VIAS TERRESTRES NO OSM .....	89
FIGURA 23 – COMBINAÇÕES ÚTEIS .....	90
FIGURA 24 – ETIQUETAS NECESSÁRIAS .....	91
FIGURA 25 – ETIQUETAS OPCIONAIS.....	92
FIGURA 26 – INSERÇÃO DAS INFORMAÇÕES SEMÂNTICAS NO OSM .....	93
FIGURA 27 – METADADOS DO CONJUNTO DE ALTERAÇÕES .....	97
FIGURA 28 – METADADOS DO CONJUNTO DE ALTERAÇÕES (cont.).....	98
FIGURA 29 – METADADOS DA LINHA RUA DES. HUGO SIMAS .....	99
FIGURA 30 – METADADOS DA LINHA RUA DES. HUGO SIMAS (cont.).....	99
FIGURA 31 – HISTÓRICO DE EDIÇÃO DA RUA DES. HUGO SIMAS.....	100

FIGURA 32 – IDENTIFICAÇÃO DO CONTRIBUIDOR .....	100
FIGURA 33 – HISTÓRICO DE EDIÇÕES DO CONTRIBUIDOR.....	101
FIGURA 34 – TRILHAS PÚBLICAS GPS DO CONTRIBUIDOR.....	102
FIGURA 35 – PROCESSO METODOLÓGICO .....	105
FIGURA 36 – MÉTODO PROPOSTO PARA O ALINHAMENTO SEMÂNTICO .....	111
FIGURA 37 – METAMODELO CONCEITUAL E SEMÂNTICO DA ET-EDGV 3.0..	132
FIGURA 38 – DIAGRAMA DO MODELO CONCEITUAL DO OSM .....	134
FIGURA 39 – CORRESPONDÊNCIA ENTRE OS ELEMENTOS CONCEITUAIS DA ET-EDGV 3.0 E DO OSM.....	136
FIGURA 40 – EXEMPLO DE TABELA DE RESULTADOS DO ALINHAMENTO SEMÂNTICO .....	142
FIGURA 41 – EXEMPLO DA TABELA DOS CASOS DE CORRESPONDÊNCIA SEMÂNTICA .....	146
FIGURA 42 – EXEMPLO DE PARTE DO GRAFO PARA O ALINHAMENTO ENTRE A CHAVE HIGHWAY DO OSM E A CATEGORIA SISTEMA DE TRANSPORTE DA ET-EDGV 3.0 .....	149
FIGURA 43 – MAPA ESQUEMÁTICO DAS CAMADAS DE FEIÇÕES NA INTEGRAÇÃO ENTRE OS MAPEAMENTOS.....	155
FIGURA 44 – MAPA ESQUEMÁTICO DAS CAMADAS DAS FEIÇÕES NA INTEGRAÇÃO ENTRE OS MAPEAMENTOS.....	157
FIGURA 45 – MAPA ESQUEMÁTICO DAS CAMADAS DAS FEIÇÕES NA INTEGRAÇÃO ENTRE OS MAPEAMENTOS.....	158

## **LISTA DE QUADROS**

QUADRO 1 – MAPEAMENTO TERRESTRE: NORMAS, PADRÕES E ESPECIFICAÇÕES.....	64
QUADRO 2 – REFLEXÕES SOBRE A QUALIDADE ENTRE OS DOIS MAPEAMENTOS .....	96



## LISTA DE TABELAS

TABELA 1 – IDENTIFICAÇÃO DE PROPÓSITO E ESPECIFICAÇÃO DE REQUISITOS .....	128
TABELA 2 – CATEGORIAS E CHAVES ALINHADAS.....	137
TABELA 3 – RESUMO DAS CATEGORIAS E CHAVES ALINHADAS.....	143
TABELA 4 – RESUMO DAS CATEGORIAS E CHAVES ALINHADAS (cont.).....	144
TABELA 5 – RESUMO DAS CATEGORIAS E CHAVES ALINHADAS (cont.).....	145
TABELA 6 – TESTE DE AVALIAÇÃO DO ALINHAMENTO CURITIBA .....	151
TABELA 7 – TESTE DE AVALIAÇÃO DO ALINHAMENTO CAMPO DOS GOYTACAZES.....	154
TABELA 8 – CLASSIFICAÇÃO EM RELAÇÃO AO NÍVEL DE DIFICULDADE DO ALINHAMENTO SEMÂNTICO .....	160
TABELA 9 – CLASSIFICAÇÃO EM RELAÇÃO AO NÍVEL DE DIFICULDADE DO ALINHAMENTO SEMÂNTICO (cont.).....	161
TABELA 10 – CLASSIFICAÇÃO EM RELAÇÃO AO NÍVEL DE DIFICULDADE DO ALINHAMENTO SEMÂNTICO (cont.).....	162
TABELA 11 – DETECÇÃO DE INCONSISTÊNCIAS NA COMPATIBILIZAÇÃO ....	163

## LISTA DE SIGLAS

AI – Artificial Intelligence  
API – Application Programming Interface  
BDG – Banco de Dados Geoespaciais  
CTB – Código de Trânsito Brasileiro  
CONCAR – Comissão Nacional de Cartografia  
COM – Component Object Model  
CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico  
CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior  
CC-BY-AS 2.0 – *Creative Commons Attribution Share Alike*  
DAML – *DARPA Agent Markup Language*  
DNER – Departamento Nacional de Estradas e Rodagem  
DNIT – Departamento Nacional de Infraestrutura Terrestre  
DOLCE – *Descriptive Ontology for Linguistic and Cognitive Engineering*  
CAD – *Computer Aided Design*  
CAD – Desenho Assistido por Computador  
DSG – Diretoria do Serviço Geográfico do Exército  
ET-ADGV 3.0 – Especificação Técnica para a Aquisição de Dados Geoespaciais Vetoriais  
ET-CQDG – Especificação Técnica para Controle de Qualidade de Dados Geoespaciais  
ET-EDGV 3.0 – Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais  
ET-PCDG – Especificação Técnica para Produtos de Conjuntos de Dados Geoespaciais  
GML – *Geography Markup Language*  
GPS – *Global Positioning System*  
HTTP – *Hypertext Transfer Protocol*  
IDE – Infraestrutura de Dados Espaciais  
INDE – Infraestrutura Nacional de Dados Espaciais  
IBGE – Instituto Brasileiro de Geografia e Estatística  
IPPUC – Instituto de Pesquisa e Planejamento Urbano de Curitiba  
IA – Inteligência Artificial  
CI – Inteligência Coletiva  
KB – *Knowledge Base*  
KB – Base de Conhecimento  
KIF – *Knowledge Interchange Format*  
LN – Linguagem Natural  
LPO – Lógica de Primeira Ordem  
CNAE – Manual de Classificação Nacional de Atividades Econômicas  
MapTopoPE – Mapeamento Topográfico Oficial em Pequenas Escalas

MapTopoGE – Mapeamento Topográfico Oficial em Grandes Escalas  
MDE – Modelos Digitais de Elevação  
OMG – *Object Management Group*  
OMT-G – *Object Modeling Technique for Geographic Applications*  
OAEI – *Ontology Alignment Evaluation Initiative*  
OIL – *Ontology Inference Layer*  
OWL – *Ontology Web Language*  
ODbL 1.0 – *Open Data Commons Open Database License*  
OGC – *Open Geospatial Consortium*  
OGC – *Open GIS Consortium*  
OSM – *OpenStreetMap*  
OS – *Ordnance Survey*  
Perfil MGB – Perfil de Metadados Geoespaciais do Brasil  
POI – Pontos de Interesse  
PLN – Processamento de Linguagem Natural  
PPGCG – Programa de Pós Graduação em Ciências Geodésicas  
RCO – Relações de Classes de Objetos  
REST – *Representational State Transfer*  
RDF – *Resource Description Framework*  
RBC – Revista Brasileira de Cartografia  
URBE – Revista Brasileira de Gestão Urbana  
SCN – Sistema Cartográfico Nacional  
SGBD – Sistema Gerenciador de Banco de Dados  
SIG – Sistemas de Informação Geoespacial ou  
SIG – Sistemas de Informação Geográfica  
SGBD – Sistemas Gerenciadores de Bancos de Dados  
KBMS – Sistemas Gerenciadores de Base de Conhecimento  
SPARQL – *SPARQL Protocol and RDF Query Language*  
SQL – *Structured Query Language*  
UML – *Unified Modeling Language*  
URIs – *Uniform Resource Identifier*  
UFPR – Universidade Federal do Paraná  
VGI – *Volunteered Geographic Information*  
XML – *Extensible Markup Language*  
W3C – *World Wide Web Consortium*

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>1</b>
1.1	HIPÓTESE .....	5
1.2	OBJETIVOS .....	6
1.2.1	Objetivo Geral.....	6
1.2.2	Objetivos Específicos .....	6
1.3	JUSTIFICATIVA.....	7
1.4	ESTRUTURA DA TESE .....	10
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA .....</b>	<b>11</b>
2.1	INTEROPERABILIDADE SEMÂNTICA DE DADOS ESPACIAIS.....	11
2.1.1	Semântica Aplicada a Dados Geoespaciais .....	16
2.1.2	Ontologias .....	18
2.1.2.1	Definição de Ontologias.....	19
2.1.2.2	Componentes das Ontologias e a Categorização dos Objetos .....	21
2.1.2.3	Tipos de Ontologias.....	33
2.1.2.4	Interoperabilidade entre Ontologias.....	37
2.1.2.5	Estudos e Aplicações .....	43
2.2	MODELAGEM CONCEITUAL DE DADOS GEOESPACIAIS.....	47
2.2.1	Níveis de Abstração de Aplicações Geoespaciais.....	48
2.2.2	Modelo OMT-G.....	50
2.3	MAPEAMENTO TOPOGRÁFICO EM GRANDES ESCALAS X MAPEAMENTO CADASTRAL NO BRASIL .....	59
2.4	MAPEAMENTO TOPOGRÁFICO OFICIAL NO BRASIL.....	59
2.4.1	Conceito de mapeamento topográfico oficial .....	60
2.4.2	Legislação, Instituições e Competências.....	61
2.4.3	Normas, Padrões e Especificações .....	62
2.4.3.1	Modelo Conceitual .....	65
2.4.3.2	Qualidade .....	72
2.4.3.3	Metadados .....	77



2.5	MAPEAMENTO COLABORATIVO .....	79
2.5.1	Conceito de mapeamento colaborativo .....	79
2.5.2	<i>OpenStreetMap</i> .....	82
2.5.2.1	Modelo Conceitual .....	85
2.5.2.2	Qualidade .....	93
2.5.2.3	Metadados .....	96
2.6	INTEGRAÇÃO DOS DADOS COLABORATIVOS AO MAPEAMENTO TOPOGRÁFICO OFICIAL .....	102
<b>3</b>	<b>METODOLOGIA.....</b>	<b>103</b>
3.1	IDENTIFICAÇÃO DE PROPÓSITO E ESPECIFICAÇÃO REQUISITOS ...	105
3.2	CAPTURE DA ONTOLOGIA .....	106
3.2.1	Obtenção das informações semânticas para o alinhamento .....	108
3.2.2	Estrutura dos modelos conceituais.....	109
3.2.3	Seleção das chaves do <i>OSM</i> e categorias da ET-EDGV 3.0 .....	109
3.2.4	Método proposto para o alinhamento semântico .....	110
3.3	FORMALIZAÇÃO DA ONTOLOGIA .....	117
3.4	INTEGRAÇÃO COM ONTOLOGIAS EXISTENTES.....	123
3.5	AValiação da qualidade da ontologia.....	123
3.6	DOCUMENTAÇÃO DA ONTOLOGIA.....	126
<b>4</b>	<b>RESULTADOS E DISCUSSÕES .....</b>	<b>127</b>
4.1	IDENTIFICAÇÃO DE PROPÓSITO E ESPECIFICAÇÃO REQUISITOS ...	127
4.2	CAPTURE DA ONTOLOGIA .....	131
4.2.1	Estrutura dos modelos conceituais.....	131
4.2.2	Seleção das chaves do <i>OSM</i> e categorias da ET-EDGV 3.0 .....	137
4.2.3	Método proposto para o alinhamento semântico .....	141
4.3	AValiação da qualidade da ontologia.....	150
<b>5</b>	<b>CONCLUSÕES .....</b>	<b>164</b>
<b>6</b>	<b>PESQUISAS FUTURAS.....</b>	<b>168</b>
<b>7</b>	<b>REFERÊNCIAS .....</b>	<b>170</b>

## 1. INTRODUÇÃO

O uso sinérgico entre diferentes fontes de dados geoespaciais pode potencializar o uso e a qualidade dessas informações não só para os usuários finais, mas igualmente para ambas as fontes (OLTEANU-RAIMOND *et al.*, 2017; HEIPKE, 2010). Essa pesquisa fala de integração dos dados do mapeamento colaborativo da plataforma *OpenStreetMap* (OSM) aos dados do mapeamento topográfico oficial brasileiro.

Em virtude do baixo investimento em cartografia no Brasil, o mapeamento topográfico oficial cobre o território de modo desigual entre as pequenas e médias escalas, além de estar desatualizado em mais de trinta anos. Além disso, o mapeamento topográfico em escalas grandes, necessário ao planejamento urbano, ainda é escasso no país (CAMBOIM, *et al.*, 2015; CARISSIMI *et al.*, 2011; PEREIRA *et al.*, 2003).

Em contrapartida, o mapeamento colaborativo de plataformas como o *OpenStreetMap* (OSM) produz e disponibiliza dados geoespaciais atualizados e abertos através da *internet*, especialmente em áreas urbanas (LUDWIG & ZIPF, 2019; ESTIMA & PAINHO, 2013). Embora sua qualidade seja heterogênea, estudos têm demonstrado que ela é satisfatória para muitos usos (TOUYA *et al.*, 2017; SIEBER & JOHNSON, 2013; CAMBOIM & SLUTER, 2009; BEARDEN, 2007; ESTES & MOONEYHAN, 1994).

Desse modo, o mapeamento colaborativo se apresenta como uma alternativa relevante para atualização e completude do mapeamento topográfico oficial, bem como para o uso associado a outras geotecnologias nas atividades de gestão urbana pelos municípios brasileiros.

Contudo, alcançar a interoperabilidade entre dados e sistemas de geoinformação não é uma tarefa trivial devido à complexidade das informações geoespaciais e pelas várias formas através das quais ela pode ser modelada (SONDHEIM *et al.*, 2015; JANOWICZ, *et al.*, 2013; LIMA, *et al.*, 2002; SMITH e MARK, 1998). Pode ocorrer incompatibilidade no nível sintático e/ou semântico.

O nível sintático diz respeito a conversão entre formatos de arquivos, mas atualmente existe consenso entre a comunidade internacional na utilização dos padrões do *Open Geospatial Consortium* (OGC) (OGC, 2020). Entretanto, a importação não se isenta de distorções nos dados porque cada sistema é concebido

de modo diferente em nível conceitual, lógico e físico, permanecendo um desafio o intercâmbio dos dados em nível semântico (CHEATHAM *et al.*, 2019; JANOWICZ *et al.*, 2013; LIMA *et al.*, 2002).

Para esse nível, Lima *et al.* (2002) afirmam que a falta de modelos conceituais comuns implica em problemas na troca de dados entre as organizações, entre eles, distorção dos dados, comprometimento da qualidade das informações, perda nas definições de atributos e até mesmo o georreferenciamento. Desse modo, a definição de um modelo conceitual, modelo de dados, *schema* ou esquema, conduz a um caminho para suplantar o desafio da interoperabilidade em nível semântico. Um modelo conceitual fornece uma descrição coerente e sistemática do conteúdo e organização de um conjunto de dados. Esse *schema* inclui a definição do significado das feições e demais componentes, primordiais para o intercâmbio semântico entre diferentes bancos de dados (SONDHEIM *et al.*, 2015; JANOWICZ *et al.*, 2013; LIMA *et al.*, 2002).

Sendo assim, para que a integração do mapeamento colaborativo ao mapeamento topográfico oficial bem como, a sua automatização seja possível, são necessárias considerações a respeito da integração e processos para compatibilizá-los. Essas considerações devem garantir que os conceitos e significados dos termos de cada uma das feições sejam definidos e compartilhados adequadamente para que não haja comprometimento da qualidade das informações. A primeira consideração é: como integrar os dados sem perder a integridade de cada uma das fontes e preservar o conhecimento do usuário? A segunda consideração é: como manter a conexão com os bancos de dados de origem?

A primeira consideração está relacionada ao elemento de avaliação da qualidade chamado de acurácia temática, o qual se refere à correta interpretação das feições e atributos, bem como à sua inclusão nas classes contidas no modelo conceitual. A segunda consideração está relacionada ao elemento de avaliação da qualidade chamado de consistência lógica, o qual diz respeito às regras lógicas que seguem uma especificação de estruturação dos dados nos níveis conceitual, físico e lógico. Sendo assim, tanto para integrar os dados sem perder a integridade de cada uma das fontes, quanto para que seja mantida a conexão com os bancos de dados de origem, é necessário realizar a compatibilização semântica com base nos modelos conceituais de ambos os mapeamentos.

A Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV 3.0), é o modelo conceitual e semântico do mapeamento de referência terrestre brasileiro (DSG, 2017), o qual é documentado através de um volume disponibilizado no formato PDF e cuja implementação do banco de dados é feita através de um *plugin* para o *software* livre QGIS. O modelo conceitual do *OpenStreetMap* é baseado na linguagem *Extensible Markup Language* (XML) e disponibilizado através de *wikis* na *internet*.

Entretanto, formalizar e associar conceitos e significados para os quais já existe um senso comum torna essa missão desafiadora (NOVACK *et al.*, 2019; SONDEIM *et al.*, 2015; JANOWICZ *et al.*, 2013; AL-BAKRI & FAIRBAIRN, 2012). Podem ocorrer variações no significado dos conceitos e das feições entre as diferentes culturas e lugares, já que as definições desses significados são resultantes da percepção e cognição humana, consenso social e estado do conhecimento. Essas variações podem também ocorrer dependendo do contexto geográfico e da situação de uso das informações (NOVACK *et al.*, 2019; JANOWICZ *et al.*, 2013).

De acordo com Novack *et al.* (2019) na tentativa de trabalhar através de “mundos de significado” para interoperabilidade de dados, as pesquisas em geontologia e geosseântica auxiliaram na formação de novos modelos para representar o mundo, no entanto, a questão fundamental da dependência geocultural ainda está por ser resolvido. A relevância dessa questão pode ser demonstrada pela realização do *workshop* “*Geographical and Cultural Aspects of Geo-Information: Issues and Solutions*” no Chipre em junho de 2019. Entre as questões em pauta estavam abordagens para lidar com aspectos geográficos e culturais em diferentes contextos de análise e propósitos de aplicação, generalização *versus* especificidade em Ciência da Geoinformação e, a relevância de diferentes geografias sociais e materiais para a Ciência da Geoinformação.

Os mesmos autores Novack *et al.* (2019), afirmam que a interação entre aspectos físicos e socioculturais produz e reproduz representações conceituais. Se as ontologias dos Sistemas de Informação Geográfica (SIG) devem espelhar essas representações, e se a dinâmica e o resultado dessa interação varia geográfica, histórica e socialmente, então as ontologias dos SIG devem também ser, se não específicas para cada lugar, tempo, grupo social e uso, suficientemente flexíveis para permitir a representação, sistematização e análise de diferentes aspectos geográficos e socioculturais. Nesse contexto, a crescente linha de pesquisa da compatibilização



de ontologias é uma fonte promissora para o alcance da interoperabilidade entre ontologias sejam elas genéricas e/ou específicas. A fusão de dados geoespaciais e o desenvolvimento de bancos de dados com incorporação de contexto são também abordagens de pesquisa que contribuem significativamente para a interoperabilidade das ontologias de SIG, o que amplia as possibilidades epistemológicas (NOVACK *et al.*, 2019).

No contexto geoespacial, um dos objetivos do uso das ontologias é conseguir melhor entendimento do universo geográfico, bem como, facilitar a troca de informações entre diferentes grupos de especialistas e usuários na manipulação das entidades geográficas pelos sistemas de informações geoespaciais, evitando distorções oriundas da cognição humana sobre os fenômenos geográficos e provendo características comuns a esses sistemas (SMITH e MARK, 2001; SMITH e MARK, 1998).

Nesse sentido cita-se pesquisas como a de Cheatham *et al.* (2019) que apresenta um alinhamento automático de ontologias de domínio de Hidrografia. Nessa pesquisa foram feitos alinhamentos semânticos manualmente para estabelecer padrões de referência para a avaliação do desempenho dos sistemas utilizados e da qualidade dos alinhamentos automatizados. Contudo, os resultados mostraram que os sistemas de alinhamento existentes não funcionaram muito bem neste domínio e que nenhum deles permitiu encontrar relações de equivalência diferentes de um para um. Esses resultados demonstram que somente as ontologias não são suficientes para compatibilizar os conceitos das diversas fontes e que os sistemas para compatibilização utilizados no estudo ainda não estão desenvolvidos suficientemente para fazer a automatização com resultados satisfatórios.

Ludwig e Zipf (2019) fizeram um estudo sobre as *tags* do *OpenStreetMap* que descrevem as áreas verdes urbanas das cidades de Munique e Dresden na Alemanha, Dar es Slaam na Tanzânia e Tel-Aviv em Israel. Os autores concluíram que a representação da vegetação em áreas urbanas na plataforma OSM é influenciada pelo contexto sócio-cultural e pelo propósito de produção do mapa. A conclusão dos autores corrobora a necessidade e a importância das pesquisas com novas abordagens para lidar com aspectos geográficos e culturais locais em diferentes contextos de análise e propósitos de aplicação.

Estima e Painho (2013) utilizaram o CORINE *Land Cover* e o *OpenStreetMap* para estabelecer as correspondências entre as nomenclaturas das duas fontes. De

acordo com os autores a acurácia global, de aproximadamente 76%, torna os resultados “notáveis e promissores”. Entretanto, a acurácia da classificação de áreas úmidas que corresponde à *tag natural:wetland* do OSM foi de apenas 1,2%. Esse resultado reforça as conclusões de Ludwig e Zipf (2019) de que pode haver um viés ocidental nas definições de algumas *tags* do OSM e que, existe dependência do contexto cultural no caso de compatibilizações semânticas como estas.

A presente pesquisa vai ao encontro das pesquisas supracitadas com foco na compatibilização semântica entre os diferentes bancos de dados. Como contribuição para essa linha de estudos apresenta-se como nova abordagem a utilização dos modelos conceituais como base para as ontologias que os integram. Sob o ponto de vista da questão dos aspectos geográficos e culturais em diferentes contextos, a presente pesquisa contribui com o exemplo brasileiro e, sob o contexto de análise e propósito de aplicação contribui com a agregação do mapeamento colaborativo do OSM ao mapeamento topográfico oficial para atualização e completude. Não se exclui a questão da generalização *versus* especificidade, pois embora seja utilizado o caso específico do Brasil, o método proposto na presente pesquisa pode ser generalizado para aplicação em outros contextos geográficos.

Deste modo, o problema desta pesquisa configura-se o seguinte, como integrar semanticamente o modelo conceitual do mapeamento colaborativo da plataforma *OpenStreetMap* (OSM) ao de outras bases de dados como é o caso do modelo conceitual do mapeamento topográfico oficial no Brasil.

## 1.1 HIPÓTESE

Se a integração for realizada em nível semântico considerando a interpretação das feições e atributos, bem como, a sua inclusão nas classes contidas nos modelos conceituais;

Se forem respeitadas as regras lógicas de estruturação dos dados nos níveis conceitual, físico e lógico em cada um dos modelos conceituais e;

Se for criada uma base de conhecimento e uma ontologia que descreva o alinhamento entre os conceitos das classes, feições e atributos dos dois modelos conceituais;

Então será possível integrar o banco de dados do mapeamento colaborativo da plataforma *OpenStreetMap* (OSM) ao banco de dados do mapeamento topográfico oficial no Brasil, preservando e permitindo conectar o conhecimento do leigo ao do especialista, do global ao local e subsidiar a geração de novas ferramentas de inteligência artificial e sistemas baseados em conhecimento.

## 1.2 OBJETIVOS

### 1.2.1 Objetivo Geral

Propor um método de compatibilização semântica entre o modelo conceitual do mapeamento colaborativo da plataforma *OpenStreetMap* (OSM) e a Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV 3.0), o modelo conceitual do mapeamento topográfico oficial brasileiro através de uma ontologia que alinha os dois modelos.

### 1.2.2 Objetivos Específicos

1) Efetuar um estudo semântico prévio para clarificar o entendimento do significado no uso dos termos mapeamento topográfico em grandes escalas e mapeamento cadastral no Brasil;

2) Realizar um teste prático de integração entre os bancos de dados do mapeamento colaborativo e do mapeamento topográfico oficial em grandes escalas de uma área de estudo, para testar as potencialidades e desafios da compatibilização semântica para atualização e completude da base oficial;

3) Desenvolver um método de correspondência semântica entre os mapeamentos, através do alinhamento das descrições dos conceitos, propriedades e relacionamentos entre os elementos constantes dos modelos conceituais de ambas as fontes;

4) Compilar diagramas dos modelos conceituais, seus conceitos e relações, bem como do alinhamento semântico entre eles e, analisar os resultados do processo de integração;

5) Gerar uma base de conhecimento e uma ontologia da compatibilização entre os modelos conceituais para ser compartilhada, reutilizada e atualizada colaborativamente;

6) Demonstrar a aplicabilidade do uso do alinhamento semântico proposto, através do *plugin* desenvolvido por (Silva *et al.*, 2020, no prelo) no contexto dessa pesquisa, o qual automatiza a integração dos bancos de dados do mapeamento colaborativo e do mapeamento topográfico oficial brasileiro, bem como avaliar a qualidade do alinhamento.

### 1.3 JUSTIFICATIVA

O mapeamento topográfico oficial é produzido e administrado por instituições públicas ou privadas designadas para esse fim, obedece a normas técnicas e a legislação específicas e, se constitui na base cartográfica imprescindível a qualquer Sistema de Informação Geográfica (SIG) (COLEMAN, 2013).

Dependendo do contexto, o mapeamento oficial pode ser disponibilizado de modo gratuito, mas, em alguns casos, existem custos e restrições ao acesso e ao uso. Os dados podem também ser protegidos por direitos autorais ou regidos por acordos formais ou licenças de uso. Além disso, parte do mapeamento oficial pode ter o acesso limitado por algumas instituições por razões de segurança, proteção dos dados ou benefícios econômicos da nação (COLEMAN, 2013).

Embora o mapeamento topográfico oficial demande alto custo e morosidade na sua produção (OLTEANU-RAIMOND *et al.*, 2017; GOODCHILD, 2009; ESTES & MOONEYHAN, 1994), por ser executado pelas instituições de governo e obedecendo aos padrões estabelecidos, tem alto grau de confiabilidade (COLEMAN, 2013).

Entretanto, devido ao baixo investimento em cartografia no país, o mapeamento topográfico oficial brasileiro cobre o território de forma desigual entre as escalas pequenas e médias escalas (1:250.000, 1:100.000, 1:50.000 e 1:25.000), além de estar desatualizado em mais de trinta anos (CAMBOIM, *et al.*, 2015). Além disso, o mapeamento topográfico em escalas grandes (maiores do que 1:25.000), necessário ao planejamento urbano, ainda é escasso no país. Ademais, a competência para sua produção é compartilhada entre a União, os Estados e os municípios e não existem legislação e normas específicas que atendam às

peculiaridades deste tipo de mapeamento no país. Apenas em 2016, foram lançadas as normas para aquisição (DSG, 2016c) e estruturação (DSG, 2016a) de dados geoespaciais englobando as escalas grandes, constituindo-se em importantes especificações para a produção dos novos mapeamentos.

Como solução à longo prazo para estes e outros problemas, foi criada através de decreto (BRASIL, 2008) a Infraestrutura Nacional de Dados Espaciais (INDE), uma iniciativa para integrar os dados geoespaciais existentes nas diversas instituições do governo brasileiro, harmonizando-os e proporcionando a sua disseminação e uso efetivo através de um geoportal. Entre os objetivos da INDE estão, evitar a duplicidade de ações e o desperdício de recursos na obtenção de dados geoespaciais pelos órgãos da administração pública e, promover a utilização dos padrões e normas homologados pela Comissão Nacional de Cartografia (CONCAR) na produção dos dados e informações geoespaciais pelos órgãos públicos de todos os níveis de governo (INDE, 2017).

Por outro lado, a emergência do *Volunteered Geographic Information* (VGI) ou mapeamento colaborativo, em plataformas como o *OpenStreetMap* (OSM), produzem e disponibilizam dados geoespaciais atualizados e abertos através da *internet*. O OSM é uma rica fonte de informações geoespaciais especialmente em áreas urbanas, incluindo informações semânticas que não são de cunho administrativo e comercial oficial (LUDWIG e ZIPF, 2019; NOVACK *et al.*, 2018; ESTIMA e PAINHO, 2013). Embora sua qualidade seja heterogênea, estudos têm demonstrado que ela é satisfatória para muitos usos (TOUYA *et al.*, 2017; SIEBER & JOHNSON, 2013; CAMBOIM & SLUTER, 2009; BEARDEN, 2007; ESTES & MOONEYHAN, 1994).

A relevância das informações geoespaciais colaborativas pode ser demonstrada ainda, pelo fato de que muitas pesquisas têm sido realizadas nos últimos anos em relação à avaliação da sua qualidade global com o intuito de utilizá-las na atualização do mapeamento topográfico oficial em vários países (RÖNNEBERG *et al.*, 2018; OLTEANU-RAIMOND, *et al.*, 2017; TOUYA, *et al.*, 2017; SIEBER & JOHNSON, 2013; ESTIMA E PAINHO, 2013; AL-BAKRI e FAIRBAIM, 2012; BEARDEN, 2007). Essas pesquisas incluem a integração do mapeamento colaborativo com o mapeamento topográfico oficial em grandes escalas (TOUYA, *et al.*, 2017; SIEBER & JOHNSON, 2013; BEARDEN, 2007) bem como, a sua utilização em países como o Brasil, nos quais este tipo de informação é praticamente inexistente (ESTES & MOONEYHAN, 1994; CAMBOIM & SLUTER, 2009).

Esses dados podem então, fornecer informações de grande valor para dar suporte ao planejamento urbano e à tomada de decisões nos municípios, pois são produzidos por moradores locais que têm conhecimento da dinâmica espacial de onde vivem e dos fenômenos que ali ocorrem (SIEBER & JOHNSON, 2013). Além disso, os dados do mapeamento colaborativo têm agregado o valor intrínseco da participação cidadã, a qual no caso do Brasil está prevista no Estatuto da Cidade (BRASIL, 2001) para a gestão democrática das cidades.

A adoção do modelo conceitual estabelecido até o momento para o Brasil, a Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV 3.0) (DSG, 2017) juntamente com outras normas e padrões, colabora para o alcance da interoperabilidade semântica na produção de informações geoespaciais nos municípios e no país como um todo. Além disso, facilita a reutilização dos dados, de modo a evitar o desperdício de recursos com a obtenção dos dados e promove ainda, a transparência nas instituições do governo (DSG, 2017).

Acrescenta-se que transparência estimula a participação popular, a qual é um elemento de destaque no Estatuto da Cidade (BRASIL, 2001). A participação da comunidade na gestão e no planejamento do território é essencial para que os problemas sejam diagnosticados mais facilmente e, conseqüentemente para que as soluções sejam efetivas. Uma destas formas de participação é através dos dados colaborativos, motivo que têm levado muitos governos a utilizá-los em conjunto com o mapeamento topográfico oficial. Uma sociedade bem informada é motivada a se comprometer com o alcance de resultados que levam à melhoria da qualidade de vida de todos. A parceria entre o planejamento técnico e os processos participativos resulta em políticas públicas de êxito e com responsabilidades compartilhadas entre os governos e a sociedade como um todo (SIEBER & JOHNSON, 2013).

É importante destacar que esta pesquisa de doutorado tem papel relevante nos temas desenvolvidos pelo grupo de pesquisa em Cartografia e SIG na UFPR, que tem trabalhado com questões do mapeamento colaborativo. A compatibilização semântica realizada nesta pesquisa de doutorado compõe a etapa inicial e imprescindível para as etapas subsequentes da integração entre os bancos de dados do mapeamento colaborativo e do mapeamento de referência, tema pesquisado por outros membros do grupo.

A pesquisa do doutorando, Caio dos Anjos Paiva trata da avaliação da qualidade intrínseca do mapeamento colaborativo e a automatização da avaliação. A

pesquisa do doutorando e analista do IBGE, Leonardo Scharth Loureiro Silva trata da integração do mapeamento colaborativo ao mapeamento de referência em pequenas escalas abrangendo o processo de avaliação de qualidade dos dados, generalização cartográfica e metadados. Em sua pesquisa de doutorado Everton Bortolini trata questões do mapeamento colaborativo para o mapeamento de favelas no Brasil. A pesquisa de mestrado de Idalécio Pascoal Joaquim mostra um panorama da utilização do mapeamento colaborativo do *OpenStreetMap* em resposta a passagem do ciclone Idai pela cidade da Beira em Moçambique. A pesquisa de doutorado de Elias Nasr Naim Elias visa o desenvolvimento de ferramentas para avaliação e visualização da qualidade do mapeamento colaborativo.

Todas essas pesquisas estão inseridas no contexto do projeto “Soluções em *software* livre para Infraestrutura de Dados Espaciais e Sistemas de Informação Geográfica”. O projeto é desenvolvido pelos professores e alunos de pós-graduação do Laboratório Geoespacial Livre da Universidade Federal do Paraná (UFPR) que faz parte da Rede Geo *for All*. Este projeto de pesquisa visa desenvolver material de treinamento, desenvolvimento de *software* e aplicações que visem a interoperabilidade de informações geoespaciais, embasem a criação de políticas públicas, permitam o treinamento de produtores e usuários de geoinformação e criem formas inovadoras para o compartilhamento, atualização e análises de dados.

#### 1.4 ESTRUTURA DA TESE

O capítulo 2 comporta as seções da revisão de literatura. Na seção 2.1 foi abordada a questão da interoperabilidade semântica e o uso de ontologias. Na seção 2.2 foram apresentados os principais conceitos referentes à modelagem conceitual de dados geoespaciais. Na seção 2.3 foi apresentado o artigo publicado na Revista Brasileira de Cartografia (RBC) no qual foi feita a distinção entre os conceitos de mapeamento topográfico em grandes escalas e de mapeamento cadastral no Brasil, estudo necessário no direcionamento do escopo da pesquisa. Na seção 2.4 foram tratados os aspectos relativos ao mapeamento topográfico oficial no Brasil, modelo conceitual, legislação e normas vigentes. Na seção 2.5 foi tratado o mapeamento colaborativo, seu conceito, o modelo conceitual utilizado, bem como a descrição dos elementos relativos à sua qualidade e metadados. Na seção 2.6 foi apresentado o



artigo publicado na Revista Brasileira de Gestão Urbana que contém um estudo de caso realizado para a qualificação com o intuito de levantar os requisitos e testar a viabilidade da integração semântica do mapeamento colaborativo ao mapeamento topográfico oficial no Brasil. Na seção 3, após a revisão teórica, foram descritas as etapas metodológicas e o seu desenvolvimento para o alinhamento semântico das ontologias dos mapeamentos. Na seção 4 foram apresentados os respectivos resultados e as discussões correspondentes. Na seção 5 foram apresentadas as conclusões da pesquisa. Na seção 6 foram feitas as sugestões para pesquisas futuras. Na seção 7 estão as referências bibliográficas.

## **2. FUNDAMENTAÇÃO TEÓRICA**

### **2.1 INTEROPERABILIDADE SEMÂNTICA DE DADOS ESPACIAIS**

Nessa primeira seção da revisão teórica são tratadas as principais questões e conceitos relativos à interoperabilidade semântica aplicada a dados geoespaciais. Inicialmente foram discutidas as peculiaridades inerentes as informações geoespaciais dentro do contexto semântico. Em seguida foram tratadas as ontologias e suas características relevantes como uma solução para a questão da interoperabilidade semântica. Posteriormente foi discutido o papel das bases de conhecimento nesse contexto. Na sequência foram tratadas as metodologias e tecnologias utilizadas na construção de ontologias e os diferentes processos de integração entre elas. Por fim, foi apresentado o estado da arte no uso de ontologias geoespaciais para interoperabilidade semântica de dados.

A interoperabilidade entre sistemas de geoinformação não é uma tarefa trivial devido à complexidade das informações geoespaciais (JANOWICZ, *et al.*, 2013; LIMA, *et al.*, 2002; SMITH e MARK, 1998). Por isso, antes de entrar na questão da interoperabilidade propriamente, é importante destacar os principais aspectos relativos à complexidade das informações geoespaciais.

As informações geoespaciais são domínio da Ciência da Geoinformação, um campo de pesquisa interdisciplinar, abrangente e ubíquo. A geoinformação está presente nas Ciências da Terra, Ciências do Mar, Geologia, Geografia, Ecologia e Arqueologia entre outras Ciências. O uso dos Sistemas de Informação Geográfica

(SIG) e as análises espaciais ampliam ainda mais esse espectro para aplicações como planejamento urbano, economia, saúde e segurança pública, por exemplo. Ademais, como é de interesse da Ciência da Geoinformação os processos que ocorrem nos fenômenos por ela analisados, acrescenta-se à dimensão espacial, a dimensão temporal, caracterizando as informações geoespaciais como espaço-temporais e aumentando ainda mais a sua complexidade (JANOWICZ *et al.*, 2013; SMITH e MARK, 1998). Além disso, com a ampla utilização das tecnologias de localização por satélite por usuários não-especialistas, é possível afirmar que a maioria das informações produzidas na atualidade tem algum tipo de referência espacial (JANOWICZ *et al.*, 2013; GOODCHILD, 2007). Nesse contexto, podem ocorrer problemas de interoperabilidade na troca de informações entre os diversos usuários e sistemas.

Os problemas de interoperabilidade têm sido alvo de preocupação e pesquisas desde o final da década de 1970. As incompatibilidades entre os sistemas de informações são agravadas pela inerente complexidade da informação geoespacial e pelas várias formas através das quais ela pode ser modelada (SONDHEIM *et al.*, 2015). As dificuldades podem ocorrer em nível sintático e/ou semântico.

Desde os anos 1990, a maioria das pesquisas sobre interoperabilidade foi concentrada no nível sintático, com foco no intercâmbio de dados geoespaciais e do desenho assistido por computador (CAD) usando tradutores bidirecionais que operam nos formatos de arquivos (SONDHEIM *et al.*, 2015). Entretanto, a importação dos dados pelos sistemas não é isenta de distorções nos dados, porque cada sistema é concebido de modo diferente em nível conceitual, lógico e físico (LIMA *et al.*, 2002).

Por esse motivo, uma segunda abordagem para o problema da interoperabilidade em nível sintático, é a utilização de um modelo conceitual universal ou genérico. Essa é a ideia do *Open GIS Consortium* (OGC) e seu objetivo principal é desenvolver uma *interface* comum, baseada em tecnologias de computação distribuída como *Extensible Markup Language* (XML), *Component Object Model* (COM), *Structured Query Language* (SQL) e JAVA, a qual é implementada seguindo o conceito de *Application Programming Interface* (API) como é o caso da ISO 19125:2004 (ISO/TC, 2004) (SONDHEIM *et al.*, 2015; LIMA *et al.*, 2002). Contudo, essa solução resolve apenas a questão da padronização no acesso a bancos de dados geoespaciais, permanecendo um desafio o efetivo intercâmbio de dados entre os

sistemas em nível semântico (CHEATHAM *et al.*, 2019; JANOWICZ *et al.*, 2013; LIMA *et al.*, 2002).

Uma primeira abordagem para o problema da interoperabilidade semântica foi o uso de metadados. Os metadados ou “dados sobre os dados” podem ser utilizados para fornecer uma descrição de alto nível sobre um conjunto de dados e incluem o formato dos dados, propriedade, linhagem, qualidade e sistema de referência, entre outros elementos (GUPTILL, 2015). Embora os metadados possam fornecer também definições concisas a respeito do conjunto de dados, bem como da sua estrutura e organização servindo de base para busca e acesso, a ênfase normalmente é no processo de produção dos dados (GOODCHILD & LONGLEY, 2015; LIMA *et al.*, 2002).

A definição de um modelo conceitual, *schema* ou esquema, nos conduz a um caminho para suplantarmos o desafio da interoperabilidade em nível semântico. Um modelo conceitual fornece uma descrição coerente e sistemática do conteúdo e organização de um conjunto de dados. Esse *schema* inclui a definição semântica das feições, o detalhamento dos nomes e tipos de atributos, um dicionário que define a estrutura dos objetos geográficos e seus atributos espaciais e não-espaciais, metadados proporcionando uma visão sinóptica das informações e um dicionário de termos relacionados (SONDHEIM *et al.*, 2015). O nível semântico se refere então, à representação conceitual das informações geoespaciais, bem como a definição do significado das feições e demais componentes dentro de cada esquema, primordiais para o intercâmbio de dados entre diferentes bancos de dados (JANOWICZ *et al.*, 2013; LIMA *et al.*, 2002).

De acordo com Sondheim *et al.* (2015), questões-chave para a interoperabilidade semântica incluem o modo como os modelos conceituais são representados, como um *schema* determinado por um usuário refere-se a conceitos geoespaciais padronizados (p.ex. sistema de referência, geometria, etc.) e como esquemas definidos por diferentes usuários se relacionam. Os padrões de estruturação de dados geoespaciais fornecem um alto grau de consistência semântica na medida em que os conceitos geoespaciais são bem definidos.

Fora os conceitos, todavia, a consistência semântica permanece um problema. A integração entre sistemas envolve a geração de um modelo conceitual comum que seja capaz de descrever a maior parte ou toda a informação de interesse. Ocasionalmente, uma alternativa mais prática é estabelecer uma série de

transformações de modelo para modelo de modo que um dos esquemas seja redefinido da maneira mais apropriada ao esquema de destino. Embora essa alternativa seja mais prática, ela só é viável com o conhecimento detalhado de cada um dos modelos conceituais e com *software* para transformação entre eles (NOVACK *et al.*, 2019).

Desse modo, conclui-se que para que interoperabilidade ocorra entre os diversos sistemas e usuários, é necessário que os conceitos e significados dos termos cada uma das feições sejam definidos e, compartilhados adequadamente. Entretanto, o fato de haver a necessidade de formalizar conceitos e significados para os quais já existe um senso comum torna essa missão desafiadora (SONDHEIM *et al.*, 2015; JANOWICZ *et al.*, 2013). Sob esse aspecto, são descritos a seguir, três exemplos que demonstram que significados de termos espaciais aparentemente óbvios não são assim tão simples e caso essa premissa não seja considerada, tarefas comuns podem ter resultados equivocados.

Deseja-se descrever o significado do termo montanha. De acordo com os autores Janowicz *et al.* (2013) (tradução livre), montanha “é um tipo de feição geográfica que estamos acostumados a utilizar em nossa comunicação diária e para a qual desenvolvemos noções internas através de nossas experiências físicas e sociais”. Como no caso de outros termos como rio, lago, floresta e cidade, com os quais estamos familiarizados, é presumido que o seu entendimento seja partilhado. Entretanto na realidade, cada um desses termos tem diferentes significados em domínios específicos do conhecimento e muitas vezes, significados incompatíveis (JANOWICZ *et al.*, 2013).

Um método para conceituar a noção de montanha é definir, como uma propriedade necessária, a sua altura mínima. No primeiro exemplo, Janowicz *et al.* (2013) dizem que essa técnica adotada no Reino Unido foi utilizada de modo bem-humorado no filme *The Englishman Who Went Up a Hill But Came Down a Mountain* (“O Inglês que Subiu a Colina e Desceu a Montanha”). No filme, a comunidade galesa ficou desapontada ao saber que a sua montanha local, era na verdade uma colina, pois tinha dois pés (aproximadamente 61 centímetros) abaixo do limiar de mil pés (aproximadamente 304 metros) de altura. No final, eles conseguem ter a sua colina classificada como uma montanha adicionando um monte de terra ao seu topo. Desse exemplo, conclui-se que (JANOWICZ *et al.*, 2013) (tradução livre):

“as definições dos tipos de feições são produtos da percepção e cognição humana, estado corrente do conhecimento e acordo social. (...) Consequentemente, as definições irão variar entre lugares e culturas e, é importante representar as definições locais apropriadamente.”

Indo adiante nessa linha de raciocínio, o segundo exemplo leva à compreensão da importância do contexto geográfico local para se entender o significado dos termos geoespaciais. Trata-se de um mapa esquemático de 1846 que mostra as principais montanhas (determinadas por sua altura) e rios (determinados por seu comprimento) da Terra. Nesse caso, a montanha mais alta da Inglaterra (912 metros) se comparada à cadeia de montanhas Himalaias aparenta ser apenas uma colina (JANOWICZ *et al.*, 2013).

O terceiro exemplo traz o fato de que os termos geoespaciais têm também seus significados dependentes da situação, já que resultam da percepção humana. Nesse exemplo, tem-se um sistema baseado em localização e alimentado através de geossementica que é projetado para fornecer instruções de localização. Sendo a montanha nesse caso, uma feição marcante na paisagem devido a sua altura, uma pessoa ao fornecer instruções de localização poderia dizer “vire a direita e caminhe em direção à montanha”. Em um sistema desse tipo, o significado de montanha não é baseado em uma definição canônica, mas totalmente baseado na situação e no contexto da rota especificada (JANOWICZ *et al.*, 2013).

Esses exemplos demonstram que muitos conceitos geoespaciais têm perspectivas múltiplas, sua interpretação depende do espaço, do tempo e às vezes de um grupo específico de usuários ou especialistas. De acordo com Sheider (2012) “o problema não é que as máquinas não consigam se comunicar, mas que os humanos não se entendem quando se comunicam através das máquinas.” Descrever formalmente a semântica dos termos geoespaciais em sistemas de informação e compartilhar seus significados ainda é um desafio e Janowicz *et al.* (2013) afirmam que (tradução livre):

“Há de se destacar que a pesquisa em geossementica não está interessada em superar ou resolver a questão da heterogeneidade semântica. Definições locais existem por bons motivos. Se alguém pudesse apenas padronizar os significados globalmente, não haveria a necessidade da pesquisa em geossementica. Em vez disso, a geossementica, tenta restringir o vocabulário de domínio ao significado pretendido... Sem melhores geo-ontologias, (meta) dados semanticamente anotados e serviços de alinhamento de ontologias mais complexos, que possam fazer o mapeamento entre ontologias

locais, reutilizar e integrar dados de fontes heterogêneas é apenas um sonho distante.”

Al Bakri e Fairbairn (2012) afirmam que a avaliação da similaridade das propriedades semânticas de objetos espaciais é uma concepção que precisa ser abordada no contexto da integração de dados em Ciência da Geoinformação na busca pela interoperabilidade entre diferentes conjuntos de dados.

Afirmações como as de Janowicz *et al.* (2013) e de Al Bakri e Fairbairn (2012) corroboram a necessidade e a relevância de alinhamentos semânticos como os realizados na presente pesquisa, que propõe um alinhamento semântico entre os dados do mapeamento topográfico oficial “local” do Brasil com os dados do mapeamento de referência colaborativo “global” do *OpenStreetMap*.

Ao longo dessa seção foram destacados os principais aspectos relativos à complexidade das informações geoespaciais e as questões de nível sintático e semântico que devem ser levadas em conta para tornar efetiva a interoperabilidade entre essas informações. Na próxima seção, foi apresentado um panorama das pesquisas semânticas e sua evolução no contexto geográfico.

### 2.1.1 Semântica Aplicada a Dados Geoespaciais

A semântica é um campo da linguística que estuda a significação das palavras bem como, a representação do sentido dos enunciados. Enquanto ciência a semântica estuda a evolução do significado das palavras, signos e símbolos utilizados na comunicação humana (MICHAELIS, 2019).

No contexto geográfico a chamada geossemântica ou semântica geoespacial tem seu foco nas expressões linguísticas referentes à Ciência da Geoinformação incluindo a *GeoWeb*. Como subárea da Ciência da Geoinformação, a geossemântica também envolve diversas áreas de pesquisa (HU, 2017; AGARWAL, 2005; MARK *et al.*, 2000). De acordo com Hu (2017) é possível destacar seis grandes áreas de pesquisa dentro da geossemântica as quais podem estar interconectadas e envolvidas simultaneamente em uma mesma pesquisa. São elas (tradução livre): interoperabilidade semântica e ontologias; *gazetteers* digitais; recuperação de informações geoespaciais; *Web* semântica geoespacial e dados

interligados (*linked data*); semântica de lugares e; cognição dos conceitos geográficos e raciocínio qualitativo.

A geossementica pode ser definida como o entendimento do conteúdo dos sistemas de geoinformação e a captura desse entendimento por teorias formais (KUHN, 2005). Desta definição destaca-se as expressões, “o entendimento do conteúdo” e “teorias formais”.

O sujeito do “entendimento do conteúdo” de um sistema de informação geoespacial pode ser o homem ou a máquina. Quando o sujeito é o homem, as pesquisas em semântica geoespacial envolvem a cognição humana em relação aos conceitos geográficos e às relações espaciais. No caso da máquina, as pesquisas envolvem a interoperabilidade semântica de sistemas distribuídos, *gazetteers* digitais e recuperação de informações geográficas. No que se refere às “teorias formais”, as ontologias têm sido a principal técnica utilizada para capturar o entendimento dos sistemas de geoinformação. As ontologias são uma forma de estruturar o conhecimento com base na descrição de conceitos e relacionamentos, que vêm sendo amplamente aplicada em geossementica (HU, 2017). As ontologias serão abordadas na próxima seção.

Com o avanço nas tecnologias da computação e da informação, e o grande fluxo de dados na *internet*, as pesquisas em geossementica também se ramificaram a *Web*. A chamada *Web Semântica* se refere a uma extensão da *Web*, na qual as palavras têm seu significado definido com o objetivo de estruturar os conteúdos publicados, para facilitar o entendimento e a colaboração entre computadores e usuários. A partir do seu desenvolvimento, foram introduzidas novas funcionalidades para relacionar os conceitos dos dados disponíveis universalmente e melhor aproveitá-los, retornando resultados relevantes às buscas dos usuários na *internet* (BERNERS-LEE, *et al.*, 2001). De acordo com seus idealizadores, Berners-Lee, Hendler e Lassila (2001), para que as pessoas e os computadores trabalhem em colaboração, é necessário que o conteúdo de dados na *Web* tenha seu significado melhor definido.

Com o *Big Data*, houve também um crescimento no volume de texto não estruturado em linguagem natural (LN) resultante de mídias sociais, *blogs* e verbetes na *Wikipedia*. Embora subjetivos, os dados textuais revelam o entendimento e a percepção dos indivíduos sobre o ambiente social, cultural, natural e geográfico. Pesquisas como a de (HU *et al.*, 2015) apresentam o uso de dados não-estruturados



para extração de semântica de lugares com o objetivo de entender padrões espaço-temporais entre pessoas e lugares. Com o desenvolvimento da inteligência artificial (IA), do processamento de linguagem natural (PLN), das técnicas de mineração de dados (*data mining*) e aprendizado de máquina (*machine learning*) - todos sistemas baseados em conhecimento (*knowledge based systems*) - , tendem a surgir novos temas de pesquisas em geossementica (HU, 2017).

De modo mais genérico, embora no passado problemas de alto custo e escalabilidade tenham desencorajado o desenvolvimento de abordagens baseadas em conhecimento, nas últimas décadas, a crescente disponibilidade de recursos de conhecimento colaborativo *online* tem tornado possível combater as dificuldades na aquisição de conhecimento por meio de uma colaboração massiva. Sob esse aspecto, os repositórios de conhecimento construídos de forma colaborativa, como a *Wikipedia*, têm sido usados como fontes de ampla cobertura de informações semiestruturadas e anotações manuais para uma ampla variedade de aplicativos de IA e PLN. (HOVY *et al.*, 2013). No contexto geográfico, a plataforma *OpenStreetMap* atua da mesma forma, produzindo informações geoespaciais e suas descrições semânticas, constituindo-se também em uma base de conhecimento geoespacial colaborativa.

Dentre as abordagens apresentadas para a aplicação da semântica aos dados geoespaciais, as ontologias têm papel relevante para o contexto dessa pesquisa, e por isso são abordadas mais detalhadamente na próxima seção.

### 2.1.2 Ontologias

Os recursos *online* são uma das principais forças motrizes por trás de abordagens que exploram grandes quantidades de conhecimento legível por máquina para executar tarefas que exigem inteligência humana. Sendo assim, esse conhecimento encontra-se no cerne da inteligência artificial e do processamento de linguagem natural (HOVY *et al.*, 2013).

Entretanto, vários problemas técnicos se interpoem no caminho dos sistemas baseados em conhecimento para compartilhamento e reutilização. Assim como nas aplicações convencionais, os sistemas baseados em conhecimento são fundamentados sobre plataformas de *hardware* heterogêneas, linguagens de



programação e protocolos de rede. Contudo, os sistemas baseados em conhecimento apresentam requisitos especiais para que se tornem interoperáveis. Tais sistemas operam e se comunicam utilizando declarações em uma representação formal do conhecimento. Eles permitem fazer consultas e fornecem respostas, tendo como *input* uma base de conhecimento (*knowledge base*) (KB) através de declarações. Além disso, como agentes em um ambiente distribuído de inteligência artificial, eles negociam e trocam conhecimento (GRUBER, 1993a).

De acordo com Gruber (1993a), para que a comunicação em nível de conhecimento ocorra são necessárias convenções em três níveis: formato de linguagem de representação, protocolo de comunicação do agente e especificação do conteúdo do conhecimento compartilhado. Propostas para padrões de formatos de representação do conhecimento e linguagens de comunicação de agentes são independentes do conteúdo de conhecimento sendo trocado ou comunicado.

Toda base de conhecimento, sistema baseado em conhecimento ou agente em nível de conhecimento está vinculado a alguma conceitualização, explícita ou implicitamente. “Uma ontologia é uma especificação explícita de uma conceitualização” (GRUBER, 1993b). Desse modo, as ontologias podem desempenhar a função de especificação de *software* para estabelecer acordos sobre conhecimento através de premissas compartilhadas e modelos do mundo.

#### 2.1.2.1 Definição de Ontologias

Borst (1997) define as ontologias como “uma especificação formal e explícita de uma conceitualização compartilhada”. Essa definição, embora sintética, traz consigo conceitos importantes. “Especificação formal” significa que é legível por computadores. “Explícita” quer dizer que os conceitos, propriedades, relações, funções, restrições e axiomas são expressamente definidos. A saber, um axioma é uma premissa considerada necessariamente evidente e verdadeira, sendo apenas uma expressão lógica formal (GRUBER, 1993a). Nos modelos de dados, os axiomas são os relacionamentos entre os elementos conceituais e as restrições de integridade espacial. “Conceitualização” se refere a um modelo abstrato de um fenômeno do mundo real. Por fim, “compartilhada” significa que existe consenso em relação ao conhecimento partilhado.

O principal objetivo para a criação de ontologias é possibilitar o compartilhamento e a reutilização de conhecimento (BORST, 1997; GRUBER, 1993a). No contexto geoespacial, de acordo com Smith e Mark (1998), as ontologias do domínio geoespacial são projetadas para produzir melhor entendimento sobre a estrutura do mundo geográfico. Para esses autores, as ontologias geoespaciais podem ter importância prática de pelo menos quatro modos.

Primeiro, podem ajudar a compreender como diferentes grupos de pessoas trocam informações geoespaciais ou falham ao trocá-las. Segundo, podem contribuir para o entendimento de determinados tipos de distorções características das relações cognitivas quanto aos fenômenos espaciais. Sob esse aspecto, de acordo com os autores, existem tendências, principalmente na conceitualização de entidades geopolíticas subjacentes a determinados tipos de conflitos territoriais. Terceiro, o estudo das ontologias sobre as entidades geográficas, sobretudo as de nível básico manipuladas nos sistemas de informação geoespacial, podem fornecer características predefinidas para tais sistemas. Quarto, os tipos de entidades geográficas são uma questão central para os padrões de intercâmbio de dados, pois uma parte substancial da semântica dos dados pode acompanhar os tipos de instâncias que integram tais entidades (SMITH e MARK, 2001; SMITH e MARK, 1998).

Em Ciência da Computação, para possibilitar o compartilhamento dos conceitos integrantes de uma ontologia, esses necessariamente precisam de uma especificação formal. Na prática, uma ontologia pode ser representada através de qualquer linguagem não formal, ou seja, através de linguagem natural, embora a representação formal seja considerada a mais apropriada (MORAIS & AMBRÓSIO, 2007). Guarino (1998) apresenta simplificadaamente, as definições formais de uma conceitualização da seguinte forma:

Seja considerada a estrutura  $\langle D, W \rangle$ . Onde,  $D$  representa um domínio e  $W$  representa os conceitos existentes em  $D$ . Tal estrutura é denominada “espaço de domínios”. Uma conceitualização  $C$  é uma estrutura  $\langle D, W, R \rangle$  na qual,  $R$  é o conjunto de relações definidas como apropriadas para representar o domínio. Sendo assim, uma conceitualização define uma estrutura projetada do mundo, representada por  $S$ . Computacionalmente, uma conceitualização necessita ainda, ser especificada em uma linguagem  $L$ .

As ontologias são instrumentos para especificação de domínios específicos do conhecimento e não do conhecimento como um todo. Quando o conhecimento é representado através de um formalismo declarativo, o conjunto de objetos passível de ser representado é chamado de “universo do discurso”. Esse universo é refletido no vocabulário representacional utilizado para representar o conhecimento em um sistema baseado em conhecimento (KBS) (GRUBER, 1993b).

Um sistema baseado em conhecimento é composto de uma base de conhecimento (KB) que representa fatos sobre o mundo e um mecanismo de inferência que pode “raciocinar” sobre esses fatos e usar regras e outras formas de lógica para deduzir novos fatos ou destacar inconsistências (HAYES-ROTH *et al.*, 1983). De acordo com Jarke *et al.* (1989) uma base de conhecimento é uma representação de informações heurísticas e factuais, frequentemente na forma de fatos, asserções e regras de dedução.

Empiricamente, uma ontologia define um conjunto de termos, que será usado para efetuar consultas. Na ontologia são definidas as regras de combinação entre os termos e seus relacionamentos, os quais são elaborados por especialistas de domínio. Os usuários, que podem ser também especialistas, por sua vez, formulam consultas utilizando os conceitos especificados na ontologia (MORAIS & AMBRÓSIO, 2007).

#### 2.1.2.2 Componentes das Ontologias e a Categorização dos Objetos

De acordo com MORAIS & AMBRÓSIO (2007) a estrutura das ontologias pode variar de acordo com os objetivos e funções propostos, mas existem alguns componentes que fazem parte da maior parte das ontologias. Os componentes básicos de uma ontologia são: (a) classes e subclasses, ou seja, o conjunto de objetos categorizados; (b) propriedades, que são as características ou qualidades das classes; (c) instâncias, que correspondem aos valores das classes e subclasses do domínio modelado; (d) regras e axiomas, que são as proposições que impõem condições e permitem realizar inferências automáticas e; (e) relacionamentos, ou seja, as ligações entre as classes (GUIZZARDI, 2005). Da análise desses componentes é possível perceber mais adiante, na seção que trata da modelagem

conceitual de dados geoespaciais, que são todos os mesmos componentes de um modelo conceitual.

No caso de ontologias do espaço geográfico e sobre os objetos e fenômenos do espaço geográfico é importante destacar alguns aspectos que as diferenciam de ontologias de outros domínios. Na maior parte das ontologias não-geoespaciais o “o que” e o “onde” são quase sempre independentes, enquanto no universo geoespacial o “o que” e o “onde” são interligados. No universo geoespacial, as categorizações são sempre dependentes de tamanho e escala, por exemplo, lago, rio, mar e oceano. Uma semelhança entre as ontologias de outros domínios e as ontologias do domínio geográfico é que as conceitualizações podem variar individual e culturalmente (SMITH e MARK, 1998a).

Em relação à categorização dos objetos (ou organização do conhecimento) em classes e subclasses, Smith e Mark (1998) afirmam que a Ciência sempre modelou as categorias de objetos como conjuntos (no sentido matemático). Um modelo teórico de um conjunto de categorias é normalmente feito através de esquemas de codificação de recursos de modo explícito (como nos padrões de dados cartográficos) ou implícito. Desse modo, assume-se que cada objeto ou evento é membro de um dado conjunto. E, sendo assim existem procedimentos ou regras para determinar a sua afiliação a cada conjunto.

Segundo Tversky & Hemenway (1984), acreditava-se inicialmente que o agrupamento dos objetos nesses conjuntos ou categorias era feito de modo arbitrário, de acordo com alguma convenção que variava entre as diferentes culturas. Entretanto, pesquisas realizadas em filosofia, antropologia, lingüística e psicologia (Berlin, 1972; Berlin, Breedlove & Raven, 1966, 1973; Brown, 1977, 1979; Hampton, 1979, 1981; Rosch & Mervis, 1975; Rosch *et al.*, 1976) identificaram semelhanças na classificação entre os diferentes idiomas e conectaram características de “categorias naturais” em relação à estrutura percebida no mundo real. Essas mesmas pesquisas possibilitaram avanços científicos e tecnológicos (Tomlinson & Petchenik, 1988; Coppock & Rhind, 1991; Montello, 2002; Lobben, 2004; Elzakker, 2004) também no que diz respeito à Geoinformação (MACEACHREN, 1995; PETERSON, 1995; FAIRBAIN, 1994).

Os procedimentos e regras para afiliação de cada objeto ou evento a determinado conjunto ou categoria, seguem uma organização que ocorre através de processos cognitivos na mente dos indivíduos. Estes processos cognitivos são

responsáveis pela abstração das informações do mundo real e pela organização destas informações em categorias (MACEACHREN, 1995; TVERSKY & HEMENWAY, 1984; ROSCH, 1973; LAKOFF, 1987).

De acordo com Rosch (1973) e Rosch *et al.* (1976) a organização das categorias em “estruturas de conhecimento” ocorre por intermédio de “processos de classificação” como a taxonomia (“por tipos”) e a partonomia (“por partes”) (os quais serão tratados mais adiante) (TVERSKY & HEMENWAY, 1984). MacEachren (1995) descreve as “estruturas de conhecimento” como modelos conectados através de ligações. Nesta visão, os conectores são as categorias produzidas mentalmente quando ocorre o processamento das informações e, as ligações entre os conectores são os relacionamentos entre estas categorias. As categorias são geradas de acordo com diferentes **níveis de abstração** e as conexões ou relacionamentos são efetuados por meio de diferentes **tipos de raciocínio** (MACEACHREN, 1995).

Rosch (1973), psicóloga e lingüista, desenvolveu a chamada Teoria do Protótipo. Segundo esta teoria a inserção de um objeto em uma categoria ocorre de acordo com a sua semelhança com um protótipo que representa o membro mais típico da categoria. Dessa teoria derivou a Teoria das Categorias de Nível Básico, segundo a qual, os indivíduos fazem uso de estratégias para armazenar mentalmente informações sobre os objetos, o que ocorre de acordo com três **níveis de abstração**, o nível superior ou superordenado (*superordinate*), o nível básico (*basic level*) e o nível inferior ou subordinado (*subordinate*). No nível superior o raciocínio é mais abstrato e no nível inferior é mais concreto (ROSCH, 1973; ROSCH *et al.*, 1976; MACEACHREN, 1995).

Para as autoras, Rosch *et al.* (1976), por “categoria” (p.ex. animal) entende-se um número de “objetos” (p. ex. cachorro, gato, etc.) que são considerados equivalentes. Por “taxonomia” entende-se um sistema através do qual, categorias se relacionam umas com as outras por meio de inclusão de classes. Por “nível de abstração” entende-se um nível particular de inclusão. Rosch *et al.* (1976) argumentam que categorias nas taxonomias de objetos concretos (em contrapartida às categorias biológicas) são estruturadas de modo que há um nível de abstração no qual os cortes de categorias mais básicos podem ser feitos: são as categorias de nível básico de abstração.

As categorias de nível básico são aquelas que carregam mais informação, possuem a mais alta “validade de sugestão” (*cue validity*) e que são mais

diferenciáveis de outras categorias. A validade de sugestão é um conceito probabilístico que pode ser definido como a probabilidade condicional de que um objeto se enquadre em uma determinada categoria, segundo uma característica ou sugestão específica. Uma alta validade de sugestão para um dado recurso significa que o recurso ou atributo possui melhor diagnose de associação à classe, do que um recurso cuja validade de sugestão é baixa (ROSCH *et al.*, 1976).

As categorias de nível superior têm baixa validade de sugestão se comparadas às categorias de nível básico, porque elas têm menos atributos comuns a outras categorias no mesmo nível (p. ex. veículos e móveis compartilham menos atributos do que carros e cadeiras). As categorias de nível inferior têm validade de sugestão menor do que as categorias de nível básico, porque compartilham também a maioria dos atributos com outras categorias de nível inferior (p. ex. cadeiras de cozinha compartilham a maioria dos seus atributos com outros tipos de cadeiras) (TVERSKY & HEMENWAY, 1984; ROSCH *et al.*, 1976).

Deste modo, existem grupos de entidades, como pássaros, por exemplo, que compartilham muitos atributos (p. ex. asas, bico, penas, pés, etc.) entre eles e que compartilham poucos atributos com outras entidades. Estas são as chamadas “categorias naturais”, ou seja, categorias que refletem a estrutura do mundo real, categorias que agrupam coisas que compartilham os mesmos atributos (TVERSKY & HEMENWAY, 1984; ROSCH *et al.*, 1976; ROSCH & MERVIS, 1975). As categorias de nível básico, ou “nível preferencial de referência” são uma característica das categorias naturais conectadas a estrutura precebida do mundo.

As categorias de nível básico são aquelas categorias cujo nível de especificidade dentro de uma taxonomia são psicológica e linguisticamente mais primárias do que categorias mais gerais (categorias de nível superior) e mais específicas (categorias de nível inferior) (TVERSKY & HEMENWAY, 1984; ROSCH *et al.*, 1976).

Conforme detectado na pesquisa de Rosch *et al.* (1976), o número de atributos listados pelos indivíduos nas categorias de nível básico, foi maior do que nas categorias de nível superior e inferior de abstração. Por exemplo, os indivíduos listaram bem poucos atributos para veículos, móveis e ferramentas, mas listaram bem mais atributos para carro, mesa e martelo. Apenas poucos atributos adicionais foram listados para carro de duas portas, mesa de jogos e martelo tipo bola (TVERSKY & HEMENWAY, 1984; ROSCH *et al.*, 1976).

Em um teste (tipo de nome dado em uma nomeação livre de objetos em fotografias), a frequência no uso de palavras de nível básico foi maior do que dos níveis superior e inferior. Neste caso, as autoras Rosch *et al.* (1976), argumentaram que, nessa tarefa de produção de nomes, os sujeitos simplesmente responderam com a palavra (frequente) mais disponível para esse objeto. Embora as frequências de palavras não tenham sido obtidas para as classes subordinadas (porque geralmente são frases, não palavras únicas), as frequências de palavras estavam disponíveis para nomes de nível básico e superior em cinco taxonomias. Em nove dos quinze casos de comparação de nível superior-básico, o nome superior realmente possuía uma frequência de palavra mais alta que o nome do nível básico. No entanto, nestes casos, os sujeitos não mostraram maior tendência a nomear com o nome superior do que nos casos em que o nome do nível básico era mais freqüente (ROSCH *et al.*, 1976).

Nas taxonomias dos nomes concretos, os objetos básicos fazem parte das categorias mais inclusivas cujos membros: (a) possuem atributos comuns a todos ou à maioria dos membros do grupo; (b) têm programas motores (*motor programs*) que são similares entre si (no caso de objetos que têm como atributo funções com movimento ou que demandam um movimento, como sentar em uma cadeira, por exemplo); (b) têm formas similares e; (d) podem ser identificados por formas semelhantes e reconhecíveis. A saber, um programa motor é uma representação abstrata do movimento que organiza e controla os possíveis movimentos necessários à execução de uma ação. Essa representação ocorre primeiramente no cérebro e antecede o movimento em si (ROSCH *et al.*, 1976).

De acordo com Tversky & Hemenway (1984) as categorias de nível básico são as categorias mais gerais com membros com formas semelhantes e reconhecíveis. Elas também são as categorias mais abstratas para as quais uma única imagem pode ser formada para a categoria. As categorias de nível básico são as categorias mais gerais, cujos membros interagem da mesma maneira (ROSCH *et al.*, 1976). Ao se rotular um objeto, são preferidos os termos de nível básico (Brown, 1958; Cruse, 1977; Rosch *et al.*, 1976) e em uma verificação, os termos básicos são verificados mais rapidamente (Murphy & Smith, 1982; Rosch *et al.*, 1976). Acrescenta-se ainda que os termos do nível básico tendem a ser as primeiras categorias nomeadas e compreendidas pelas crianças (Mervis & Rosch, 1981), os primeiros termos a entrarem em um léxico (Berlim, 1972; Rosch *et al.*, 1976), os



mais curtos e menos derivados e contextualmente neutros (Cruse, 1977). No que diz respeito ao desenvolvimento motor sensorial, os objetos básicos devem ser as primeiras categorias aprendidas por meio da percepção visual e da interação motora sensorial com o objeto (TVERSKY & HEMENWAY, 1984; ROSCH *et al.*, 1976).

Nos estudos de Rosch (1978) foi sugerido que as categorias de nível básico são mais informativas porque, dado o nosso sistema cognitivo e a estrutura do mundo real, este é o nível onde as correlações naturais e descontinuidades entre as feições são mais salientes. Deste modo, presume-se que no nível básico de abstração a informatividade é maior e consequentemente, é também mais elevado o potencial de inferência na categorização (TVERSKY & HEMENWAY, 1984). Nos estudos de Rosch (1976, 1978), é evidenciado o fato de que os indivíduos utilizam as categorias de nível básico como pressuposto para categorizar os objetos em estruturas que possam também ser compreendidas por outras pessoas (TVERSKY & HEMENWAY, 1984).

Conforme citado anteriormente, para a estruturação do conhecimento, além dos diferentes **níveis de abstração** utilizados na construção das categorias, são efetuadas também conexões ou relacionamentos entre as categorias por meio de diferentes **tipos de raciocínio** (MACEACHREN, 1995). De acordo com Brown (1958), o raciocínio humano é efetivado através de dois mecanismos, em função da capacidade de abstração individual. No primeiro mecanismo o raciocínio humano utiliza as categorias de nível básico e inferior, o qual é chamado de mente “concreta”. No segundo mecanismo o raciocínio humano utiliza as categorias de nível superior, e é chamado de mente “abstrata”. Nesta mesma linha de pensamento, Lloyd *et al.* (1996) afirmam que o nível superior de abstração é estimulado pelo raciocínio abstrato, enquanto o nível inferior é estimulado pelo raciocínio concreto (TVERSKY & HEMENWAY, 1984; BROWN, 1958).

Foi com base nos conceitos de Brown (1958), Lloyd *et al.* (1996) e Rosch (1976, 1973), que Tversky & Hemenway (1984) indicaram que a organização das categorias em estruturas hierárquicas de conhecimento ocorre por intermédio dos processos de classificação taxonomia e partonomia.

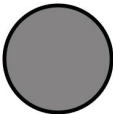
De acordo com as autoras, Tversky & Hemenway (1984), a partonomia serve para separar os objetos em seus componentes estruturais (partes) e para organizar o conhecimento da função destes objetos em relação aos componentes estruturais. No processamento das informações através da partonomia são



identificadas as partes que compõem os objetos e em seguida os objetos são agrupados de acordo com a semelhança a uma categoria prototípica. Trata-se de uma análise horizontal. Em contrapartida, as taxonomias servem para organizar numerosas classes de entidades e possibilitar inferências sobre grandes conjuntos e de conjuntos neles inclusos. No processamento das informações através da taxonomia, os conjuntos de categorias são organizados através da análise de suas partes e de outros atributos, para em seguida, serem agrupadas de acordo com a sua semelhança, por tipos. Essa tipificação pode ocorrer em diferentes níveis hierárquicos da categoria prototípica. Trata-se de uma análise vertical (TVERSKY & HEMENWAY, 1984; MACEACHREN, 1995).

Com base nos estudos de Brown (1958), Lloyd *et al.* (1996), Rosch (1976) e Tversky & Hemenway (1984), é possível inferir que a organização das categorias em estruturas de conhecimento através da taxonomia e da partonomia ocorre por meio de uma análise horizontal e posteriormente vertical ascendente (SANTIL & SLUTER, 2012). Ou seja, inicialmente são analisadas as características individuais dos objetos (estrutura, forma e função) para posteriormente agrupá-las por semelhança e categorizá-las. Neste sentido, a partonomia tende a operar através do tipo de raciocínio mais concreto e a taxonomia tende a operar através do tipo de raciocínio mais abstrato (TVERSKY & HEMENWAY, 1984). Embora o tipo de raciocínio mais abstrato esteja relacionado com o nível de abstração superior e o tipo de raciocínio mais concreto com o nível de abstração inferior, a classificação das estruturas de conhecimento pela taxonomia ou pela partonomia, independe do nível de abstração (BRAVO, 2014). A figura 1 sintetiza os conceitos envolvidos na categorização dos objetos.

FIGURA 1 – RELACIONAMENTO ENTRE NÍVEL DE ABSTRAÇÃO, OBJETO &amp; CATEGORIA E TIPO DE RACIOCÍNIO

RELACIONAMENTO ENTRE NÍVEL DE ABSTRAÇÃO, OBJETO & CATEGORIA E TIPO DE RACIOCÍNIO			
NÍVEL DE ABSTRAÇÃO	OBJETO	CATEGORIA	TIPO DE RACIOCÍNIO
Superior		Sistema de Transporte Rodoviário	Abstrato
Básico		Via	
Inferior		Trecho de Entroncamento	Concreto

FONTE: Adaptado de Bravo (2014).

Na construção de ontologias geoespaciais os processos de classificação mais usados são a taxonomia (“é um”, “tipo de”), a partonomia (“parte de”), a mereologia (teoria “parte-todo”), a cronologia (precedência entre os conceitos) e a topologia (teoria de limite e fronteira) (PINHO & GOLTZ, 2003). Os objetos geoespaciais estão intrinsecamente vinculados ao espaço de modo que herdam muitas das propriedades estruturais do espaço geográfico através dos

relacionamentos mereológicos e topológicos e da sua geometria. Devido à complexidade dos dados geoespaciais, esses relacionamentos são muito importantes para a caracterização do domínio geográfico e, uma ontologia geoespacial deve necessariamente contê-los.

Além de conter partes, os objetos geográficos também possuem limites, os quais contribuem tanto para sua composição ontológica quanto para os elementos compreendidos em seus interiores. Embora os limites não sejam objeto específico de estudo nessa pesquisa, são aqui estudados, pois têm influência na definição semântica dos conceitos geográficos. Os limites dos objetos em si são um fenômeno para o propósito de categorização e esses limites podem ser nítidos ou gradativos. Além disso, o que um objeto é, pode influenciar na localização e na estrutura dos seus limites. Se uma feição é identificada como um pântano, então seus limites estarão localizados mais distantes de uma encosta, por exemplo, do que se a feição for identificada como um rio. As entidades geográficas podem ser conectadas ou contíguas, ou ainda, distribuídas ou separadas. Às vezes fechadas, como lagos e, outras vezes abertas, como baías. Desse modo, uma ontologia do domínio geoespacial deve conter uma teoria de limites e interiores e, de conexões e separações, aliada às relações topológicas (SMITH e MARK, 2001 e 1998).

Além disso, de acordo com Smith e Mark (1998), a estrutura topológica traz consigo certos tipos de dualidade como a distinção entre um limite externo e um limite interno, ou um limite circundante de um objeto geográfico. Outra questão relacionada à noção de limites, é que ela pode significar, em diferentes contextos, um limite matemático abstrato (uma linha, plano ou superfície infinitamente fina que está localizada no espaço, mas não ocupa espaço) ou uma zona limite, de espessura pequena, mas finita. Segundo os autores, Smith e Mark (1998) os limites abstratos são vistos normalmente, como pertencentes à esfera da topologia matemática padrão. Entretanto, no domínio geoespacial, existem determinados tipos de limites com os quais a topologia padrão não consegue lidar. Se, por exemplo, um terreno ou uma ilha forem cortados ao meio, o resultado não será uma parte fechada e outra aberta. Isso ocorre porque os limites abstratos não ocupam espaço e, desse modo, podem ser perfeitamente coincidentes. Para esse tipo de fenômeno, é necessário utilizar teorias mereotopológicas especiais.

Os objetos geográficos sempre serão identificados pela especificação da localização dos seus limites. Existem dois tipos de limites, os *bona fide* e os *fiat*. Os

limites *bona fide* pertencem a objetos que possuem uma delimitação física vista de modo aproximadamente consensual entre diferentes culturas. Como exemplo, cita-se um planeta, uma estrada, um rio. Os limites *fiat* são abstratos, projetados no espaço geográfico e resultantes da cognição humana. Como exemplo, cita-se as linhas dos trópicos, os limites territoriais, os limites cadastrais e os distritos postais (SMITH e MARK, 1998).

Smith e Mark (1998), argumentam ainda, que é possível distinguir os limites *fiat* de cinco maneiras. Alguns limites podem ser vagos ou indeterminados, resultando em um problema conceitual. É o caso de uma baía representada e identificada através de toponímia em um mapa. É fácil para um observador enxergar onde está a baía. O limite exterior da baía que fica na direção do mar pode ter algum tipo de regulamentação territorial, mas essa não é importante para o contexto do exemplo. O problema é como dizer ao computador que a baía está em tal localização, que se estende de um determinado ponto a outro na linha de costa, mas que desaparece gradativamente dentro do mar e, solicitar ao computador que processe essas informações logicamente e faça inferências sobre elas como uma pessoa faria. O mesmo pode ocorrer com limites de um cabo no lado continental ou os limites do sopé de uma montanha.

A segunda forma de distinguir os limites *fiat* é como consensuais. Embora, os objetos geográficos façam parte da realidade física do espaço, são passíveis de categorizações influenciadas pela linguagem e práticas culturais de demarcação resultantes de um consenso informal. Os limites *fiat* também podem ser diferenciados como legais, aqueles impostos por leis ou normas. Nesse caso, limites de nações, setores censitários e distritos postais são como entidades sociais, com direitos, deveres e obrigações semelhantes. Inclui-se nesse tipo de limites, os definidos legalmente com base no nível médio do mar, níveis de maré, linha de costa (SMITH e MARK, 1998).

E por fim, os limites podem ser definidos cientificamente e utilizados nos Sistemas de Informação Geoespacial. Os solos são extensões contínuas de material depositado na superfície da terra. Para classificar os solos, os cientistas impõem sobre essas extensões contínuas, um conjunto de limites *fiat* determinados matematicamente através da medição de propriedades físicas e químicas de amostras. Se o sistema de classificação for alterado, os limites também o serão, alguns podem desaparecer e outros novos poderão surgir. Os polígonos assim

determinados através de formas irregulares podem distorcer o fenômeno real, mas são a melhor representação possível no estágio tecnológico atual (SMITH e MARK, 1998a).

Dentre as propriedades, as ontologias do domínio geoespacial devem levar em conta, além da geometria dos objetos, as teorias da geometria qualitativa como concavidade, convexidade, achatamento e irregularidade. Uma teoria das categorias geoespaciais deve levar em conta também que seus objetos podem ter de zero a três dimensões. Por exemplo, o Pólo Norte, o Equador, a Noruega (objetos em duas dimensões com uma curvatura em três dimensões) ou o Mar do Norte. Este último pode ser referido a um corpo água em três dimensões ou a sua superfície bidimensional. Outro exemplo, uma baía, pode se referir à porção de terra ao redor do mar ou a reentrância na linha de costa, ou a uma parte da costa, ou ainda ao corpo água. As preposições espaciais também podem tornar ambíguas as relações geométricas, por exemplo, a ilha está “dentro” do lago, em um sentido pode significar que a ilha se projeta da superfície do lago e em outro, pode significar que a ilha está completamente submersa no volume do lago (SMITH e MARK, 1998).

Para Smith e Mark (1998) a categorização das entidades no domínio geoespacial, depende de três níveis de realidade: a realidade espacial, a esfera cognitiva e a realidade geográfica. A realidade espacial, aquela que compreende todos os objetos visíveis na superfície da Terra, do mundo assim chamado real. As regiões nesse espaço real formam um sistema relacional, que inclui relações de contenção, de separação, de adjacência e sobreposição, e também, relações espaciais de localização “entre” (*between*), “sobre” (*at*) ou “em” (*in*) elementos geográficos. A esfera cognitiva é o domínio da cognição, dos conceitos, da percepção, da memória, do raciocínio e da ação. As relações existentes na realidade espacial também existem na esfera cognitiva conceitual. As representações cognitivas estão diretamente relacionadas à realidade espacial e são atreladas a cada indivíduo tanto quanto os objetos o são à realidade geográfica. Por esse motivo, conceitos como “lá” ou “aqui” são ditos egocêntricos. Os conceitos funcionam espacialmente de diversas maneiras: através de modelos abstratos ou representações do espaço na mente dos indivíduos, como quando pensamos se o “Peru está ao Norte ou ao Sul de Ghana”; quando estamos em algum lugar e utilizamos conceitos indexicais como acolá, à direita, para o leste; através da combinação dos dois anteriores; e no caso dos SIG, através da interação conceitual

com as entidades espaciais por meio de modelos matemáticos e representações computacionais.

Sob este aspecto, citam-se as três categorias de conhecimento de Golledge & Stimson (1987): conhecimento declarativo, procedimental e configuracional. Estas categorias têm ênfase na sequência de aquisição do conhecimento espacial (MACEACHREN, 1991).

O conhecimento declarativo (*declarative knowledge*) é referido pelos autores como conhecimento sobre os objetos e lugares, os quais são coisas que podem ser diretamente percebidas no ambiente. Este conhecimento inclui os atributos espaciais que permitem que os objetos e lugares sejam reconhecidos e os atributos não espaciais que permitem que o lugar seja caracterizado (MACEACHREN, 1991).

O conhecimento procedimental (*procedural knowledge*) é aquele requerido para se movimentar em um ambiente, com ênfase no processo de navegação. Neste caso, assume-se que não apenas objetos lineares são armazenados, mas também uma sequência de decisões sobre como ir de um lugar a outro. O conhecimento procedimental requer um nível de abstração mais alto do que o do conhecimento declarativo (MACEACHREN, 1991).

O conhecimento configuracional (*configurational knowledge*) está no nível mais alto de processamento cognitivo. É neste nível que o entendimento dos relacionamentos espaciais ocorre. O conhecimento configuracional inclui saber que um estado ou cidade fica em um determinado país ou, qual a posição de um determinado local em relação a outro. Além disso, o conhecimento neste nível permite que padrões geográficos sejam identificados, que relacionamentos entre padrões sejam percebidos e que hipóteses sobre associações sejam desenvolvidas (MACEACHREN, 1991).

De acordo com Thorndyke & Hayes-Roth (1982) as informações obtidas a partir de um mapa são armazenadas como imagens na mente, as quais podem ser examinadas e medidas como em um mapa físico. Essas imagens podem incluir tanto o conhecimento declarativo, quanto o procedimental. Os autores argumentam também, que experiências no mundo real produzem o conhecimento configuracional, que é armazenado como proposições que podem ser processadas para a solução de problemas (MACEACHREN, 1991).

Acrescenta-se a essa complexidade, o fato de as representações cognitivas do espaço poderem estar subdefinidas ou erradas. Podem ocorrer diferenças

individuais ou relacionadas à cultura. Os conceitos podem ser refinados ou modificados por meio de interações sociais e culturais, educação formal e definições de dicionário. Alguns conceitos espaciais podem até estar embutidos nos sistemas perceptivos dos sentidos e no cérebro. Outros conceitos podem ser alterados através do uso; assim, uma vez que um determinado conceito é considerado como correspondente a uma determinada situação, detalhes específicos dessa situação podem modificar o conceito (SMITH e MARK, 1998).

As informações supracitadas evidenciam que os indivíduos pensam e raciocinam manipulando conceitos. Os programas de computador são baseados em conceitualizações matemáticas formais dos relacionamentos entre os objetos no mundo e os indivíduos utilizam os computadores para aprender, entender ou tomar decisões sobre tais objetos. Conclui-se que o estabelecimento de correspondências e inter-relações entre os diferentes domínios de entidades e relações espaciais é essencial para a construção de Sistemas de Informação Geoespacial e de outros sistemas para raciocinar sobre entidades e relações espaciais (JANOWICZ *et al.*, 2013; SMITH e MARK, 1998).

Smith e Mark (1998) recomendam que as pesquisas e ontologias do domínio geoespacial considerem os três diferentes níveis de realidade, a do mundo real, a das representações computacionais e matemáticas e, a do campo cognitivo do raciocínio, da linguagem e das ações humanas. Os autores também sugerem que pesquisas envolvendo comparações formais de padrões de dados geoespaciais e definições de dicionário em uma variedade de idiomas, podem fornecer um ponto de partida importante para experimentos transculturais necessários para refinar os conceitos geográficos e solucionar questões de interoperabilidade semântica.

#### 2.1.2.3 Tipos de Ontologias

Segundo Almeida & Bax (2003) é possível classificar as ontologias de acordo com o grau de formalismo, tipo de aplicação, conteúdo ou função da ontologia.

Em relação ao grau de formalismo as ontologias podem ser: (a) altamente informais, quando expressas em linguagem natural; (b) semi-informais, quando expressas em uma forma restrita e estruturada de linguagem natural, aumentando

consideravelmente a clareza e reduzindo a ambiguidade; (c) semi-formais, quando expressas em linguagem computacional definida formalmente; e (d) rigorosamente formais, quando os termos são definidos através de semântica formal, teoremas e provas.

Quanto à aplicação as ontologias podem ser: (a) de autoria neutra, quando um aplicativo é descrito em uma única língua e depois convertido para o uso em outros sistemas, reutilizando-se as informações; (b) de especificação, quando é baseada em uma ontologia de domínio que é usada para documentação e manutenção no desenvolvimento de *software*; e (c) de acesso comum à informação, quando um vocabulário é inacessível e a ontologia torna a informação inteligível, conferindo conhecimento compartilhado dos termos (ALMEIDA & BAX, 2003).

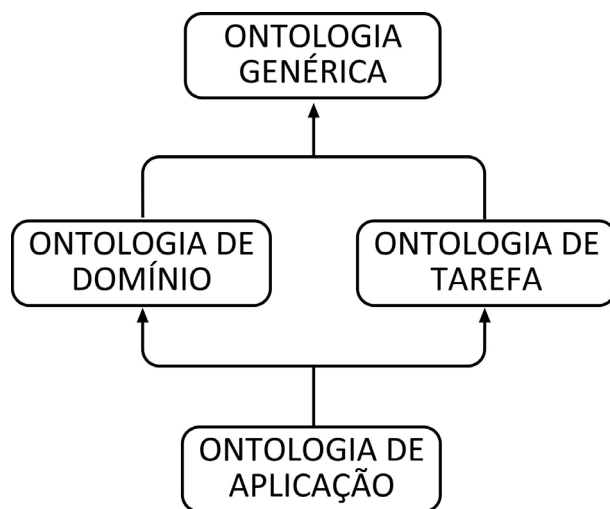
Quanto à classificação em relação ao conteúdo as ontologias podem ser: (a) **terminológicas**, quando especificam termos que serão utilizados para modelar o conhecimento de um domínio específico, como no caso dos léxicos; (b) **de informação**, quando especificam a estrutura de registros de um banco de dados; (c) de modelagem de conhecimento, quando especificam as conceitualizações do conhecimento; (d) **de aplicação**, quando contém as definições necessárias para modelar conhecimento em uma aplicação; (e) de domínio, que expressam conceitualizações específicas de um determinado domínio; (f) genéricas, que definem conceitos genéricos e comuns a diversas áreas do conhecimento; e (g) de representação, aquelas que explicam as conceitualizações que estão por trás dos formalismos de representação do conhecimento (ALMEIDA & BAX, 2003).

Quanto à função, classificação mais importante, as ontologias podem ser categorizadas de cinco formas. As ontologias (1) Genéricas ou de Topo constroem teorias básicas do mundo, de caráter abstrato e aplicáveis a qualquer domínio (de senso comum). As ontologias (2) de Domínio descrevem conceitos e vocabulários relativos a domínios específicos. Esse é o tipo mais comum de ontologia. As ontologias (3) de Tarefas expressam conceitualizações sobre a resolução de problemas utilizando o vocabulário relacionado a uma atividade ou tarefa genérica. As ontologias (4) de Aplicação descrevem conceitos que dependem de domínio e de tarefa específicos. Tais conceitos normalmente se referem a papéis desempenhados por entidades do domínio quando da execução de determinada tarefa. A figura 2 ilustra a inserção das ontologias de aplicação no contexto dos demais tipos de ontologias. As ontologias (5) de Representação explicam as conceitualizações que



fundamentam os formalismos de representação de conhecimento (ALMEIDA & BAX, 2003; GUARINO, 1998).

FIGURA 2 – TIPOS DE ONTOLOGIAS



FONTE: Adaptado de Guizzardi (2005).

No contexto geoespacial, os Sistemas de Informação Geográfica baseados em ontologias espaciais são criados para reduzir problemas de interoperabilidade e fazem uso das ontologias de aplicação (FONSECA *et al.*, 2000). De acordo com Guarino (1998), pode ser importante esclarecer a diferença entre uma ontologia de aplicação e uma base de conhecimento (tradução livre):

A explicação está relacionada ao propósito da ontologia, que é uma base de conhecimento particular, descrevendo fatos assumidos como sempre verdadeiros por uma comunidade de usuários, em virtude do significado acordado do vocabulário utilizado. Uma base de conhecimento genérica, em vez disso, também pode descrever fatos e afirmações relacionados a um estado particular das coisas (situação) ou a um estado epistêmico particular. Dentro de uma base de conhecimento genérica, é possível distinguir, portanto, dois componentes: a ontologia (contendo informações independentes de estado) e a base de conhecimento "básica" (contendo informações dependentes do estado) (GUARINO, 1998).

Uma base de conhecimento é uma forma de armazenar informações complexas estruturadas e/ou não estruturadas utilizadas por um sistema de informações. O uso inicial do termo foi relacionado a sistemas especialistas, que foram os primeiros sistemas baseados em conhecimento. A representação ideal

para uma base de conhecimento (KB) é um modelo conceitual – geralmente chamado diretamente de ontologia na literatura de inteligência artificial – com classes, subclasses e instâncias (HAYES-ROTH *et al.*, 1983). Nesse sentido uma base de conhecimento pode ser um modelo conceitual ou um banco de dados.

Indo além nessa linha de raciocínio, Hayes-Roth *et al.* (1983) afirmam que a base de conhecimento poderia ser composta por uma base de conhecimento fundamental ou básica, acrescida de uma ontologia. Para sistemas comerciais, por exemplo, poderia haver motivos para manter o conhecimento “estratégico” de forma não explícita, por razões de segurança. Neste caso, ao menos o “compromisso ontológico” ou objetivo do sistema deve ser explicitado, a fim de facilitar sua acessibilidade, manutenção e integrabilidade. Sob esse aspecto, as ontologias podem ajudar a aumentar a transparência de uma aplicação. A representação do conhecimento deve ser vista, pelo menos externamente, de modo que, a base de conhecimento sirva como uma documentação compreensível dentro e fora do grupo de desenvolvimento (JARKE *et al.*, 1989). O “compromisso ontológico”, a saber e *a grosso modo*, significa que uma teoria expressa através de “linguagem de primeira ordem está comprometida com as entidades que devem estar no domínio de interpretação para que a fórmula em questão seja verdadeira, e que essas entidades devem ser ‘dotadas de identidade’ ” (ARENHART & KRAUSE, 2009).

Sob um viés mais prático, Brachman e Levesque (1996) definem o conteúdo de uma base de conhecimento como:

“...o conjunto de possibilidades não descartadas. Se a base de conhecimento estiver vazia, tudo será possível. Uma operação chamada TELL informa a base de conhecimento sobre certas restrições no conjunto de possibilidades. A operação ASK permite ao usuário consultar a base de conhecimento sobre seu conhecimento. É mostrado que os bancos de dados são bases de conhecimento especiais nas quais apenas operações TELL muito simples são possíveis. No modelo de banco de dados relacional, por exemplo, apenas predicados n-ários livres de função e variável de estrutura predefinida podem ser afirmados. Devido a essa forma restrita da operação TELL, a operação ASK do Sistema Gerenciador de Banco de Dados (DBMS) pode ser implementada com eficiência por recuperação direta. À medida que operações TELL mais complexas (por exemplo, regras, informações incompletas, objetos complexos, etc.) são permitidas, a eficiência (com mais expressividade e até mesmo decidibilidade) é reduzida.”

Com base nessas informações, para o contexto dessa pesquisa, os modelos conceituais do mapeamento topográfico oficial e do mapeamento colaborativo são considerados ontologias. Na próxima seção é tratada a interoperabilidade entre ontologias que leva a definir o produto gerado por essa pesquisa como um alinhamento entre ontologias, o qual produz uma terceira ontologia. Essa ontologia, por sua vez, associada a toda a documentação do seu processo de geração se constitui em uma base de conhecimento que permite conectar o conhecimento do leigo ao do especialista e do global ao local para subsidiar a geração de novas ferramentas de inteligência artificial e sistemas baseados em conhecimento.

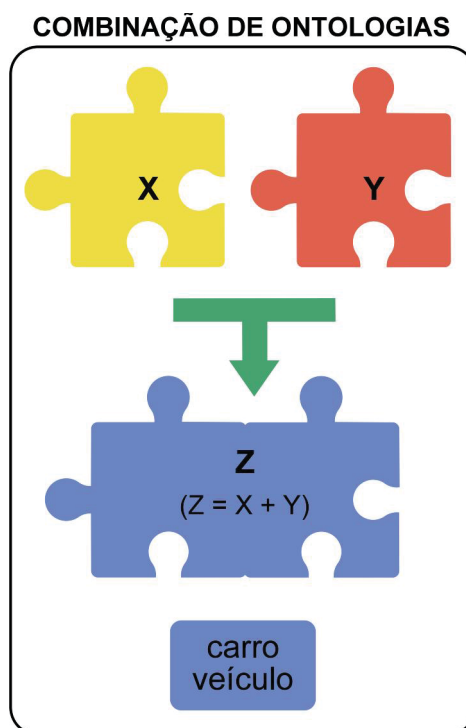
#### 2.1.2.4 Interoperabilidade entre Ontologias

A utilização de ontologias fornece uma representação comum para a troca de informações de modo que os diferentes agentes de *software* que operam nesse meio possam obter suporte no processamento automático de informações. Porém para que haja sinergia entre esses agentes na troca das informações representadas pelas das ontologias são necessários, em muitos casos, instrumentos para compatibilizá-las e assim garantir a interoperabilidade entre elas (CHEATHAM *et al.*, 2019; FELICÍCIMO, 2004).

De acordo com Noy e Musen (1999) existem duas abordagens possíveis: fundir as ontologias para gerar uma única ontologia ou alinhar as ontologias estabelecendo ligações entre elas, de modo que as ontologias alinhadas possam reutilizar as informações uma da outra. Dessas duas abordagens derivam mais duas possibilidades: combinar e mapear as ontologias. Essas quatro abordagens são explicadas a seguir.

A combinação de ontologias resulta em uma única ontologia sem a definição clara de suas origens. É comum que a combinação ocorra entre ontologias de domínios similares ou que se sobrepõem. Dois domínios se sobrepõem quando, por exemplo, um domínio X além de possuir as mesmas informações de um domínio Y, é suplementado com outras informações (NOY E MUSEN, 1999). A figura 3 ilustra a combinação de ontologias, na qual os dois conceitos similares, carro da ontologia X e veículo da ontologia Y são unidos na ontologia única Z.

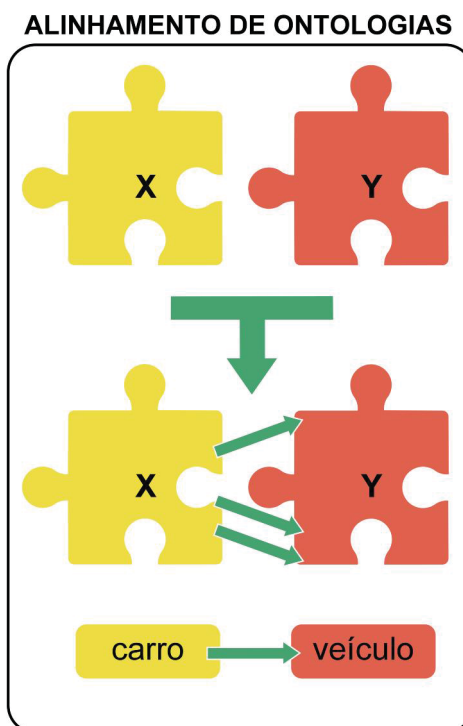
FIGURA 3 – COMBINAÇÃO DE ONTOLOGIAS



FONTE: Adaptado de Felicícimo (2004).

No alinhamento de ontologias, as ontologias são mantidas separadas de modo original, às quais são acrescentadas ligações entre seus conceitos equivalentes. O alinhamento pode ocorrer entre ontologias de domínios complementares e também de níveis diferentes. Dois domínios são complementares quando, por exemplo, um domínio Y, adiciona informações a um domínio X. Embora sejam de domínios diferentes, existem partes comuns que tratam de um mesmo assunto (NOY e MUSEN, 1999). Um exemplo é o alinhamento do domínio de *fast food* no qual só existem bebidas não alcoólicas com o domínio de bebidas que tem bebidas alcoólicas e não-alcoólicas, somente as informações comuns serão compartilhadas entre elas (FELICÍCIMO, 2004). A figura 4 ilustra o alinhamento de ontologias, no qual o conceito carro da ontologia X é alinhado ao conceito veículo da ontologia Y.

FIGURA 4 – ALINHAMENTO DE ONTOLOGIAS



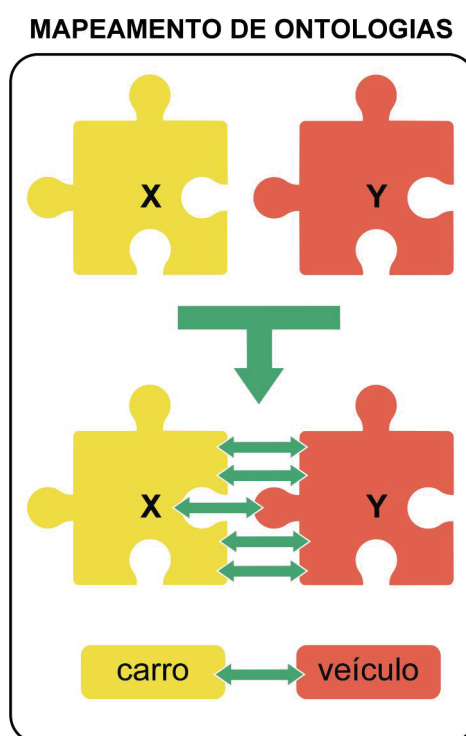
FONTE: Adaptado de Felícimo (2004).

De acordo com Cheatham *et al.* (2019), alinhamentos entre ontologias podem ser utilizados para fazer buscas nos conjuntos de dados de acordo com o vocabulário existente em ambas as ontologias, para fazer consultas de pesquisa, para executar algoritmos de inferência lógica sobre diversos domínios e outras tarefas importantes. Embora algumas dessas aplicações exijam alinhamentos de alta qualidade, os quais devem ser criados manualmente e que normalmente levam semanas mesmo para pequenas ontologias, outros usos podem se beneficiar de alinhamento automatizado sacrificando alguma precisão em detrimento de resultados oportunos. Na visão das autoras, para uso no mundo real, os sistemas de alinhamento de ontologias terão muitas vezes como resultados, ontologias indiscutivelmente imperfeitas. Sob esse aspecto, Ahn & Dabbish (2006) argumentam que as capacidades da máquina são limitadas em contraste as capacidades de raciocínio humanos.

O mapeamento de ontologias resulta em uma estrutura formal com expressões que conectam os termos entre as ontologias mapeadas. Esse procedimento pode ser utilizado para transferir instâncias de dados, esquemas de

integração e de combinação, além de outras tarefas semelhantes (FELICÍCIMO, 2004; NOY E MUSEN, 1999). Noy e Musen (1999) citam um exemplo de aplicação desse procedimento, um sistema de tradução entre idiomas. De acordo com os autores, diferenças subjacentes aos idiomas nas duas ontologias normalmente não permitem uma correspondência dos termos de um para um. Através do mapeamento de ontologias essas diferenças devem ser resolvidas. A figura 5 ilustra o mapeamento de ontologias, na qual o conceito carro de X e o conceito veículo de Y são mapeados por expressões formais.

FIGURA 5 – MAPEAMENTO DE ONTOLOGIAS

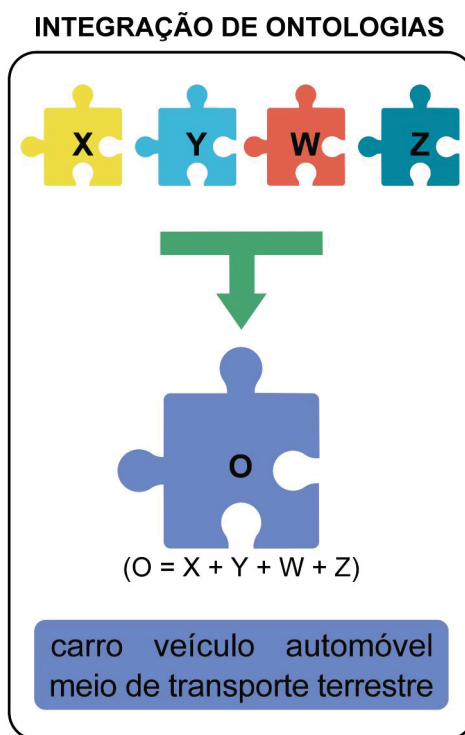


FONTE: Adaptado de Felícimo (2004).

Alguns autores fazem distinção entre o “mapeamento” e o “alinhamento” entre ontologias (FELICÍCIMO, 2004; PINTO *et al.*, 1999), enquanto outros os consideram o mesmo procedimento e utilizam os termos como sinônimos (BOUQUET *et al.*, 2003; NOY E MUSEN, 1999). De acordo com as informações apresentadas, entende-se que nessa pesquisa foi realizado um “mapeamento” entre ontologias, mas como o termo mapeamento já é utilizado primariamente no contexto da Geoinformação, optou-se por utilizar o termo “alinhamento” entre ontologias.

A integração de ontologias produz uma única ontologia resultante de diferentes operações entre elas. De acordo com Pinto *et al.* (1999), a integração de ontologias ocorre quando se cria uma única ontologia, reutilizando-se outras já existentes através de montagem, extensão, especialização ou adaptação de diferentes ontologias. Nesse procedimento são identificadas as partes originais das ontologias utilizadas para a integração. A figura 6 ilustra a integração de ontologias, onde os conceitos carro de X, veículo de Y, automóvel de W e meio de transporte terrestre de Z foram integrados na ontologia O.

FIGURA 6 – INTEGRAÇÃO DE ONTOLOGIAS



FONTE: Adaptado de Felícimo (2004).

A diferença entre a integração e a combinação é que, na combinação as ontologias combinadas são de domínios semelhantes e na integração as ontologias integradas podem ser de domínios diferentes.

De acordo Felícimo (2004) é necessário considerar alguns aspectos para minimizar os possíveis problemas nos processos de compatibilização entre as ontologias, como a presença de inconsistências nas comparações dos termos das ontologias, falta de completude no alinhamento de todos os possíveis termos e, risco

de inviabilidade da estratégia frente às inconsistências existentes. Para o caso de evitar possíveis inconsistências no processo de compatibilização a autora sugere que, além da verificação dos requisitos da compatibilização, sejam listadas as diferenças e as similaridades entre os termos, bem como a falta de informações entre eles.

A identificação de cada aspecto listado na tabela tem o propósito de refinar a qualidade do resultado da compatibilização, mensurando assim o nível de inconsistência entre as ontologias (FELICÍCIMO, 2004). Após essa identificação é necessário decidir como tratar as inconsistências. Nuseibeh *et al.* (2000) sugerem como estratégias ignorar, tolerar ou resolver as inconsistências. Como é necessário realizar a compatibilização entre as ontologias com confiabilidade, não seria uma boa solução ignorar totalmente as inconsistências. No entanto, resolver todas as inconsistências detectadas também não seria uma boa estratégia porque isso tenderia a ampliar a complexidade da solução. Conclui-se então que a permanência de algumas inconsistências é tolerável definindo-se um percentual de aceitação pré-definido.

A seleção dos recursos empregados na compatibilização deve equilibrar suas contribuições e impactos. A compatibilização total pode ser almejada, não obstante, desnecessária. É preferível garantir uma margem de alinhamento confiável, com alguns termos não alinhados, porém com baixo percentual de erro na identificação entre eles com um tempo de execução razoável (FELICÍCIMO, 2004).

Segundo, Felicícimo (2004) a compatibilização pode ser efetuada de modo automático, semi-automático, com a necessidade de intervenção humana em algumas fases para a tomada de decisão, ou ainda manualmente. No caso da solução ser implementada na *Web*, devido ao grande número de páginas e à sua natureza dinâmica, é necessário que requisitos de interoperabilidade sejam satisfeitos. Ou seja, que a solução seja feita de modo automatizado, rápido (efetuado em poucos segundos) e confiável (margem de erro mínimo na compatibilização) (HAENDCHEN *et al.*, 2003).

No caso da *internet* não há a necessidade de armazenamento dos resultados provenientes da compatibilização, ou seja, a solução pode ser não persistente, pois eles serão feitos dinamicamente. Todavia, se a compatibilização for utilizada para aplicações de inteligência artificial nas quais técnicas de aprendizado podem ser aplicadas ou em sistemas baseados em conhecimento, a solução deve



ser persistente, ou seja, seus resultados devem ser armazenados (DOAN *et al.*, 2003).

Na próxima seção serão apresentadas algumas aplicações que demonstram a relevância da utilização da semântica e das ontologias no contexto geoespacial.

#### 2.1.2.5 Estudos e Aplicações

Pesquisas como as de Smith e Mark (1998; 2001) apresentam os aspectos que diferenciam as geo-ontologias das ontologias de outros domínios no intuito de identificar as componentes essenciais para a construção de qualquer sistema de informação geográfica. Além do foco na conceitualização das componentes essenciais do domínio geográfico esse tipo de pesquisa também visa à solução de problemas de interoperabilidade.

Embora existam poucas pesquisas na literatura que propõem metodologias para desenvolvimento de ontologias, estudos como os de Hagedorn *et al.* (2019) propõem uma metodologia para desenvolvimento de ontologias de domínio para projetos de engenharia baseada em ontologias de topo. De acordo com os autores o método serve para demonstrar como a estrutura ontológica pode facilitar o projeto de produtos inovadores baseados em conhecimento, controle de projeto automatizado e com acesso baseado em consultas. No Brasil, Lima *et al.* (2002) propuseram uso de DAML (*DARPA Agent Markup Language*) associado a OIL (*Ontology Inference Layer*) como forma de descrever ontologias para os dados espaciais e promover a interoperabilidade semântica.

Outra vertente das pesquisas semânticas é a da geração de ontologias de objetos, subdomínios ou categorias geográficas específicas. No Brasil, cita-se a ontologia de referência de domínio da Divisão Político-Administrativa brasileira e seus limites geográficos que contempla as dimensões espacial, descritiva e temporal do quadro territorial brasileiro de Castañeda Filho (2017). Nessa linha, cita-se também o trabalho de Pinho & Goltz (2003) para construção de uma ontologia espacial de lote urbano. Essa representação contempla as dimensões legal, cadastral, tributária e jurídica para compatibilização dos diversos bancos de dados da Prefeitura de Belo Horizonte (MG) para elaboração de um Cadastro Técnico

Municipal informatizado. Codescu *et al.* (2011) desenvolveram uma ontologia para as *tags* do *OpenStreetMap* em língua inglesa.

Devido à importância para o contexto desta tese, são apresentadas diversas pesquisas na linha de compatibilização semântica. Cheatham *et al.* (2019) apresentam um alinhamento automático de ontologias de domínio de Hidrografia no âmbito da *Ontology Alignment Evaluation Initiative* (OAEI), uma iniciativa para comparar o desempenho de sistemas de alinhamento em tarefas padrão. Nesta pesquisa foram feitos alinhamentos semânticos manualmente para estabelecer padrões de referência para a avaliação do desempenho dos sistemas utilizados e da qualidade dos alinhamentos automatizados. Foram detectadas características comuns as ontologias utilizadas como, falta de similaridade na sintaxe dos rótulos das entidades, diferenças na granularidade da modelagem e, a tendência para as águas superficiais serem definidas em termos de outros recursos. Além disso, os resultados mostraram que os sistemas de alinhamento existentes não funcionaram muito bem neste domínio e que nenhum deles permitiu encontrar relações de equivalência diferentes de um para um.

Ludwig & Zipf (2019) mostram um estudo sobre as *tags* do *OpenStreetMap* que descrevem as áreas verdes urbanas das cidades de Munique e Dresden na Alemanha, Dar es Salaam na Tanzânia e Tel-Aviv em Israel. Através da análise dos dados com base em métodos gráficos e estatísticos os autores concluíram que a representação da vegetação em áreas urbanas na plataforma OSM é influenciada pelo contexto sócio-cultural e pelo propósito de produção do mapa.

Novack *et al.* (2018) sugerem duas estratégias de correspondência baseadas em grafos para a conjugação de pontos de interesse (POI) de dados geoespaciais das plataformas colaborativas OSM e *Foursquare*. Segundo essa estratégia os nós dos grafos representam os POIs e as extremidades representam as possibilidades de correspondência entre eles. Foram utilizadas as seguintes medidas de similaridade: espacial, que se refere a distância espacial entre os POIs de cada conjunto de dados; a similaridade entre os nomes ou *strings*, que associa o nome do ponto de interesse aos serviço, produto ou tipo de ponto; e a similaridade semântica que pode se basear em modelos conceituais, *corpus* léxico e híbrido. De acordo com os autores, os resultados atestaram que o peso ponderado da soma, utilizado nas medidas de similaridade conduziu a melhor acurácia do que quando as similaridades são somadas com pesos iguais. Os autores demonstraram ainda que

os grafos permitiram fazer correspondências de um para um, de um para muitos e de muitas para um.

Estima & Painho (2013) utilizaram o CORINE *Land Cover* e o *OpenStreetMap* para estabelecer as correspondências entre as nomenclaturas das duas fontes. O CORINE *Land Cover* é um banco de dados geoespacial temático de ocupação e uso do solo produzido pela Direção-Geral do Território, agência de mapeamento oficial de Portugal e foi utilizado como referência para avaliar a classificação das feições de área e a distribuição espacial das classes do OSM. De acordo com os autores a acurácia global, de aproximadamente 76%, torna os resultados “notáveis e promissores”, os quais encorajam as pesquisas futuras neste tópico. A acurácia da classificação das superfícies artificiais, que engloba as vias, unidades comerciais e industriais, áreas verdes, de esporte e lazer, entre outras foi de 84,3% e a da classificação de corpos de água de 99,5%. A acurácia da classificação de áreas úmidas que corresponde à *tag natural:wetland* do OSM foi de apenas 1,2%, reforçando as conclusões de Ludwig e Zipf (2019) de que pode haver um viés ocidental nas definições de algumas tags do OSM e que existe dependência do contexto cultural no caso de compatibilizações semânticas como estas.

Com o intento de avaliar uma possível integração, Al-Bakri e Fairbairn (2012), analisaram as similaridades estrutural e semântica entre os dados do *Ordnance Survey* (OS) e do *OpenStreetMap*. A similaridade semântica foi avaliada usando o pacote *WordNet:Similarity* e a similaridade estrutural entre as classificações foram modeladas através de uma árvore rotulada em esquema XML. Para os autores, os resultados revelaram que o uso de um sistema automático e genérico para realizar a correspondência semântica como o *WordNet* levou a uma má concordância entre a semântica de conjuntos de dados do OS e do OSM.

Baglatzi, *et al.* (2012) propuseram um alinhamento das *tags* do *OpenStreetMap* com a ontologia de topo *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE). Para manter o processo *bottom-up* de etiquetagem do OSM, foi utilizada a abordagem “*games with a purpose*” (jogos com um propósito, tradução livre) para auxiliar os usuários não-especialistas a alinhar suas *tags* ao DOLCE. Esse tipo de proposta tem a intenção de melhor estruturar o modelo conceitual do OSM.

No Brasil, Felicícimo (2004) desenvolveu uma estratégia de interoperabilidade que alinha automaticamente taxonomias de ontologias

comparadas. O alinhamento é realizado em três etapas seqüenciais, na primeira os conceitos das ontologias são comparados lexicalmente, na segunda as hierarquias das ontologias são comparadas estruturalmente, e na terceira etapa são refinados os resultados da etapa anterior e os conceitos são classificados, de acordo com um percentual de similaridade pré-definido.

Bouquet *et al.* (2003) propõem um algoritmo para mapeamentos semânticos cruzando classificações hierárquicas. Essa pesquisa levou o problema de coordenação semântica da lingüística computacional para o problema de dedução de relações entre conjuntos de fórmulas lógicas, as quais representam o significado dos conceitos dos diferentes modelos. No algoritmo são combinadas as informações de nível do conhecimento léxico, de domínio e de conhecimento estrutural.

Ainda na linha de compatibilização semântica, Noy e Musen (2001) e Noy e Musen (2003) apresentam o conjunto de ferramentas *PROMPT*. A ferramenta semi-automática *iPROMPT* foi projetada para realizar a combinação de ontologias interativamente, guiando o usuário durante o processo e identificando inconsistências e potenciais problemas, além de sugerir estratégias para solucioná-los. A ferramenta automática *AnchorPROMPT* foi projetada para realizar o alinhamento de ontologias com base em grafos através da identificação automática de termos semanticamente semelhantes. A ferramenta *PROMPTFactor*, faz a extração de partes de ontologias extensas para a geração de ontologias que reutilizam partes de outras ontologias. A ferramenta *PROMPTDiff* identifica diferenças em versões de uma ontologia comparando-as semântica e estruturalmente.

Os estudos descritos apresentam um painel do que vem sendo pesquisado nas duas últimas décadas quanto à aplicação das ontologias como uma solução para o problema da interoperabilidade semântica. Essas pesquisas demonstram ainda, a relevância do tema da integração do mapeamento colaborativo aos mapeamentos oficiais.

Na próxima seção, é tratada a modelagem conceitual dos dados, base para a construção da ontologia produzida nessa pesquisa.

## 2.2 MODELAGEM CONCEITUAL DE DADOS GEOESPACIAIS

Para introduzir as informações geoespaciais do mundo real, sejam eles do mapeamento topográfico, colaborativo ou a respeito de quaisquer temas, dentro de um sistema computacional é necessário ter um modelo conceitual de dados. Um modelo conceitual de dados é um conjunto de conceitos utilizados para descrever a estrutura e as operações de um banco de dados. A função do modelo é sistematizar o entendimento de cada objeto ou fenômeno a ser representado dentro do banco de dados. Contudo, os sistemas gerenciadores de bancos de dados (SGBD) têm limitações, então os objetos e fenômenos reais são simplificados e representados de modo a atender as funções do banco. Este processo de simplificação é chamado de abstração e no caso dos dados geoespaciais envolve diversos fatores (BORGES, *et al.*, 2005).

O primeiro deles é que a realidade é modelada através de conceitos geométricos. A maneira como as pessoas percebem o espaço pode ser representada com diferentes primitivas gráficas. Um objeto pode ser representado de diferentes formas dependendo da necessidade específica de cada análise ou da escala a ser utilizada. Por exemplo, uma escola, pode ser representada através de um ponto, de uma área ou, de várias áreas caso seja necessário maior nível de detalhe (DAVIS JR. & LAENDER, 2000).

O segundo fator é a inerente complexidade da informação geoespacial, conforme já discutido no capítulo anterior. Ela envolve geometria, localização no espaço, informações temporais, diferentes informações associadas e ainda, origens distintas. Quanto à origem, os objetos ou fenômenos do mundo real podem ser representados de duas maneiras: através de geo-objetos ou através de geo-campos (GOODCHILD, 1992). Alguns deles podem estar em ambas as categorias dependendo do nível de detalhamento considerado. Quando os objetos do mundo real podem ser individualizados, como é o caso de montanhas e bacias hidrográficas, por exemplo, eles são representados no banco de dados através de geo-objetos. Quando os objetos do mundo real são contínuos no espaço, como altitudes e temperaturas, por exemplo, eles são representados no banco de dados através de geo-campos (DAVIS JR. & LAENDER, 2000).

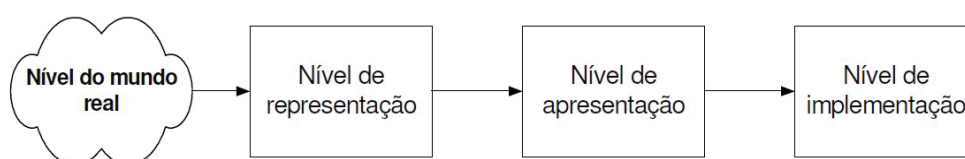
O terceiro fator é que existem relações espaciais entre os objetos dentro do banco de dados. Estas relações permitem compreender como os objetos se relacionam no mundo real (DAVIS JR. & LAENDER, 2000).

O processo de abstração dos dados geoespaciais ocorre em níveis e os modelos de dados conceituais são classificados de acordo com os níveis utilizados (BORGES, *et al.*, 2005). Na próxima seção foram apresentados os níveis de abstração.

### 2.2.1 Níveis de Abstração de Aplicações Geoespaciais

Como é possível observar na figura 7, existem três níveis de abstração empregados em aplicações geoespaciais.

FIGURA 7 – NÍVEIS DE ABSTRAÇÃO DE APLICAÇÕES GEOESPACIAIS



FONTE: Borges, *et. al* (2005).

O mundo real comporta os fenômenos geográficos a serem representados. Por exemplo: rios, cobertura vegetal, ruas.

O nível de representação comporta a modelagem dos fenômenos do mundo real através de conceitos semânticos, os quais são descritos através de anotações textuais com o significado atribuído a cada dado geoespacial. Neste nível são definidas as classes que irão compor o banco de dados (DAVIS JR. & LAENDER, 2000). Seguindo os exemplos do mundo real, poderíamos ter uma classe chamada hidrografia para conter os rios; classes como vegetação natural e vegetação cultivada para conter a cobertura vegetal e, dentro de cada uma delas possivelmente outras subclasses como mangue, cerrado, etc.; uma classe transportes, que conteria as ruas, que por sua vez seriam subclassificadas também pelo tipo.

O nível de apresentação comporta a especificação dos diferentes aspectos visuais que as entidades geográficas podem assumir durante seu uso nas aplicações geoespaciais. Neste nível são incorporadas suas características espaciais, como

localização, topologia e forma geométrica (DAVIS JR. & LAENDER, 2000). Por exemplo, um aeroporto pode ser representado por um ponto, uma área ou várias áreas dependendo da escala e da aplicação a ser utilizada.

No nível de implementação são estabelecidos os padrões, as formas de armazenamento e as estruturas de dados para cada tipo de representação. Neste nível também são definidos os relacionamentos entre as estruturas e as classes, bem como as funções e métodos necessários ao funcionamento do banco de dados espacial. Nesta fase, essas questões são tratadas computacionalmente (DAVIS JR. & LAENDER, 2000). É também nesse nível que ocorre a representação de conceitos e dinâmicas geoespaciais como ontologia computacional (NOVACK et al., 2019).

Com os níveis de abstração empregados nas aplicações geoespaciais em mente, convém aqui um parêntese para abordar a questão da escala dentro dos Sistemas de Informação Geoespacial (SIG) e da sua componente de armazenamento de dados, os Bancos de Dados Geográficos (BDG).

Os fenômenos do mundo real, modelados no nível de representação através de conceitos semânticos que descrevem o significado das feições e classes de feições, não sofrem alteração do seu significado quando assumem, no nível de apresentação, diferentes aspectos visuais durante seu uso nas aplicações geoespaciais.

A utilização do conceito de escala em SIG e BDG abrange quatro aspectos: integração de dados, visualização dos dados, análise dos dados e transformações da base de dados (D'ALGE, J. C. L. & GOODCHILD, 1996). O procedimento de aproximação ou afastamento na tela do computador, por exemplo, próprio da cartografia automatizada, causa alteração na escala de visualização dos dados (SOARES, D. & D'ALGE, 2005). Entretanto, não altera o significado das feições e classes de feições. Do mesmo modo, a existência das feições e classes de feições não se altera entre as escalas do mapeamento topográfico, pois estas refletem os elementos do mundo real. É a Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais (ET-ADGV 3.0) (DSG, 2018) que detalha os padrões para construção da geometria de cada classe de objetos, assim como os atributos necessários à individualização das instâncias de cada uma das classes por escala.

Atualmente, a estrutura da Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV 3.0) (DSG, 2017), está subdividida em



Mapeamento Topográfico em Grandes Escalas (MapTopoGE) e Mapeamento Topográfico em Pequenas Escalas (MapTopoPE), repetindo feições e classes de feições em função da abrangência das escalas. Isto ocorre porque originalmente, desde o decreto nº 243 (BRASIL, 1967), a normatização oficial do mapeamento topográfico terrestre foi feita apenas para as escalas ditas pequenas (1:25.000 e menores). As primeiras ET-EDGVs, até a versão 2.1.3 (DSG, 2010), eram feitas assim. Com o advento dos mapeamentos especiais demandados para os grandes eventos esportivos ocorridos no país entre 2013 e 2016, foi elaborada a ET-EDGV Defesa Força Terrestre (ET-EDGV F Ter) (DSG, 2016), na qual os elementos das escalas grandes foram adicionados, mas em subdivisões separadas, não tendo sido refeita (o que seria recomendável), a modelagem novamente por categorias, independente das escalas.

Esta questão será oportunamente abordada novamente quando for tratada a ET-EDGV 3.0 na seção 2.4.31 e nos resultados desta pesquisa.

Na próxima seção é apresentado o modelo OMT-G, que atua nos três níveis de abstração. Este modelo é relevante por ser utilizado na concepção da Especificação Técnica para a Estruturação dos Dados Geoespaciais Vetoriais 3.0 (ET-EDGV 3.0), o modelo conceitual brasileiro, que será tratado mais adiante e o qual foi empregado na proposta de alinhamento semântico com o mapeamento colaborativo do *OpenStreetMap*.

## 2.2.2 Modelo OMT-G

De acordo com Davis Jr. & Laender (2000) o modelo *Object Modeling Technique for Geographic Applications* (OMT-G) se utiliza das primitivas definidas para o diagrama de classes da *Unified Modeling Language* (UML), com a introdução das primitivas geográficas para ampliar a capacidade de representação semântica do modelo e reduzir a distância entre o modelo mental do espaço a ser modelado e o modelo de representação.

Através de suas primitivas, este modelo permite construir o esquema estático de aplicações geoespaciais, no qual são especificadas as classes envolvidas no problema, bem como suas representações e seus relacionamentos. A partir do esquema, podem ser produzidas as restrições de integridade espacial, as



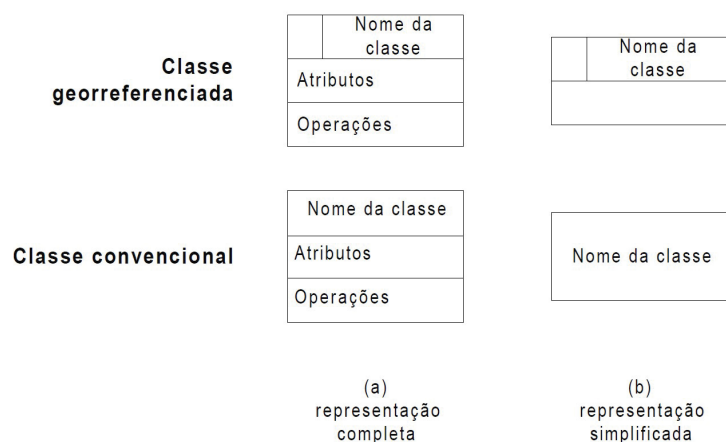
quais necessitam ser implementadas pela aplicação ou pelo banco de dados geoespaciais. A função das restrições é manter a consistência do banco de dados através de imposições (DAVIS JR. & LAENDER, 2000). Um exemplo de restrição seria: o Código Florestal atual (Lei nº 12.651/12) (BRASIL, 2012) estabelece como Áreas de Preservação Permanente (APPs) as áreas no entorno dos lagos e lagoas naturais em faixa com largura mínima de trinta metros em zonas urbanas.

As restrições de integridade podem ser restrições de domínio, de chave, de integridade referencial, de integridade topológica e de integridade semântica. De acordo com Borges *et al.* (2005) as restrições de domínio, de chave, de integridade referencial consistem “na identificação de condições que precisam ser garantidas para que o banco de dados esteja sempre íntegro”. Restrições de integridade topológica consideram as propriedades geométricas e as relações espaciais entre os objetos, como no exemplo das APPs. As de integridade semântica se referem ao significado implícito das feições geográficas. Como exemplo, uma regra que impeça que edifícios sejam interceptados por trechos de arruamento (BORGES *et al.*, 2005). Sob o aspecto das ontologias são esses os elementos que correspondem aos axiomas, conforme explicado na página 19.

O modelo OMT-G tem como base três conceitos principais: classes, relacionamentos e restrições de integridade espaciais. O conceito de restrições de integridade espaciais foi descrito nos dois parágrafos anteriores e depende das funções projetadas para cada banco individualmente (BORGES *et al.*, 2005).

As classes no modelo OMT-G podem ser georreferenciadas ou convencionais. Uma classe convencional é composta de objetos com propriedades, comportamentos, relacionamentos e semântica semelhantes, a qual pode se relacionar com objetos espaciais. A classe convencional não tem propriedades geoespaciais ou geométricas. Uma classe georreferenciada é composta de objetos que têm representação espacial, à qual são associados elementos do mundo real com suas localizações. A figura 8 mostra a notação gráfica utilizada para as classes no modelo OMT-G (DAVIS JR. & LAENDER, 2000).

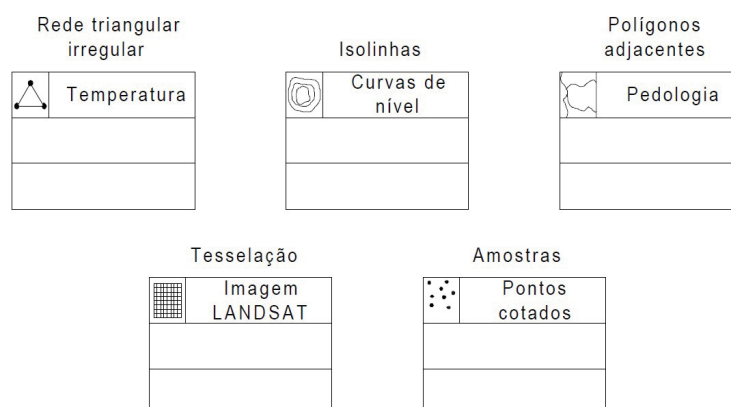
FIGURA 8 – NOTAÇÃO GRÁFICA PARA AS CLASSES DO MODELO OMT-G



FONTE: Davis Jr. &amp; Laender (2000).

A simbolização de uma classe convencional segue a notação da UML, com o nome da classe, os atributos e as operações realizadas pela classe. A notação de uma classe georreferenciada apresenta como diferença, um retângulo no canto superior esquerdo, no qual é representada a geometria utilizada. As classes georreferenciadas podem assumir a visão de campos e objetos (conforme descrito na seção 2.2). As figuras 9 e 10 mostram as geometrias possíveis para as classes georreferenciadas com distinção de acordo com geo-campos e geo-objetos (DAVIS JR. & LAENDER, 2000).

FIGURA 9 – NOTAÇÃO GRÁFICA PARA GEO-CAMPOS

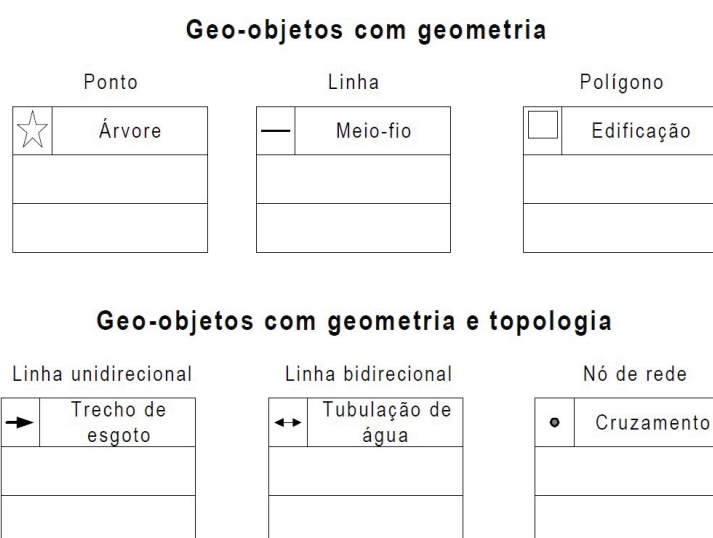


FONTE: Davis Jr. &amp; Laender (2000).

No caso dos geo-campos podem existir as seguintes classes que dela são descendentes: isolinhas, polígonos adjacentes, tesselação, amostragem e rede triangular irregular objetos (DAVIS JR. & LAENDER, 2000). No caso da ET-EDGV 3.0 só foram utilizadas isolinhas e amostragem.

No caso dos geo-objetos podem existir: geo-objetos com geometria, as quais podem ser ponto, linha e polígono; e geo-objetos com geometria e topologia, as quais podem se nó de rede, arco unidirecional e arco bidirecional objetos (DAVIS JR. & LAENDER, 2000).

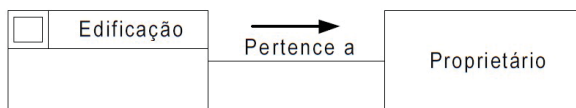
FIGURA 10 – NOTAÇÃO GRÁFICA PARA GEO-OBJETOS



FONTE: Davis Jr. & Laender (2000).

Quanto ao conceito de relacionamento, muito importante para a modelagem do espaço geográfico, o modelo OMT-G representa três tipos possíveis: associações simples, relacionamentos espaciais e relacionamentos topológicos em rede, os quais são ilustrados pela figura 11. Além destes, são possíveis mais quatro tipos derivados, a generalização, a especialização, a agregação e a generalização cartográfica, os quais foram descritos na sequência dos três primeiros objetos (DAVIS JR. & LAENDER, 2000).

FIGURA 11 – RELACIONAMENTOS



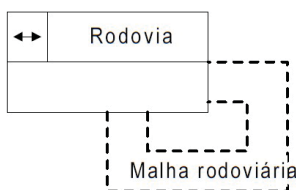
(a) Associação simples



(b) Relacionamento espacial



(c) Relacionamento de rede arco-nó



(d) Relacionamento de rede arco-arco

FONTE: Davis Jr. & Laender (2000).

A associação simples configura relacionamentos estruturais entre objetos de classes diferentes, convencionais ou georreferenciadas e é representada com linhas contínuas (figura 11a) (DAVIS JR. & LAENDER, 2000).

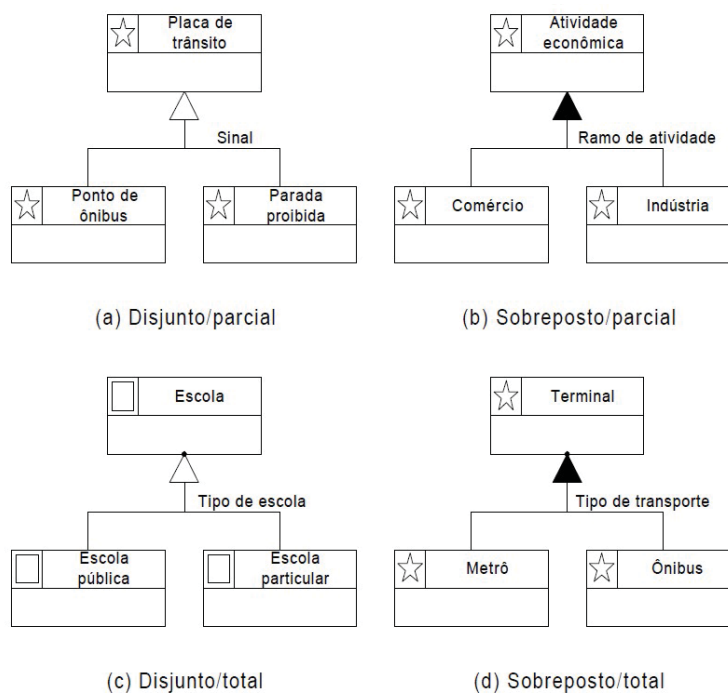
Os relacionamentos espaciais exprimem relações topológicas, métricas, ordinais e *fuzzy* e são representados com linhas pontilhadas (figura 11b). A diferenciação na representação gráfica das linhas facilita a distinção visual entre os relacionamentos. Tanto na associação simples, quanto nos relacionamentos espaciais o nome do relacionamento é anotado sobre a linha e uma seta é utilizada para indicar a direção de leitura (BORGES *et al.*, 2005; DAVIS JR. & LAENDER, 2000).

Já os relacionamentos topológicos em rede traduzem relacionamentos entre objetos conectados entre si, os quais são representados por duas linhas paralelas pontilhadas, com o nome do relacionamento anotado entre elas (figura 11c). Os

relacionamentos topológicos em rede, geralmente são especificados entre uma classe de nós e uma classe de arcos, mas estruturas de redes sem nós também podem ser definidas, gerando um relacionamento recursivo sobre uma classe de arcos, como mostra a figura 11d (DAVIS JR. & LAENDER, 2000). A ET-EDGV 3.0 não utiliza os relacionamentos topológicos em rede.

O próximo tipo de relacionamento é a generalização espacial, a qual ocorre no banco de dados. Nele, ocorre a definição de classes mais genéricas (superclasses) em função de classes semelhantes (subclasses). No relacionamento seguinte, a especialização, ocorre o processo inverso: as classes mais específicas são detalhadas a partir de uma classe mais genérica, acrescentando-se novas propriedades na forma de atributos. Neste caso, cada uma das subclasses herda atributos, operações e associações da superclasse. Os relacionamentos de generalização espacial e especialização se aplicam tanto a classes georreferenciadas quanto a classes convencionais. A notação gráfica para a generalização e a especialização é um triângulo que conecta a superclasse às suas subclasses. Pode ser indicada a propriedade ou característica abstraída pelo relacionamento de generalização espacial, como é possível observar na figura 12 (sinal, ramo de atividade, etc.) (DAVIS JR. & LAENDER, 2000).

FIGURA 12 – GENERALIZAÇÃO ESPACIAL



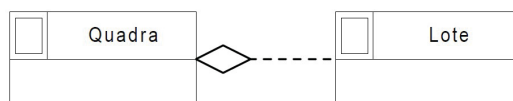
FONTE: Davis Jr. & Laender (2000).

O modelo OMT-G prevê quatro elementos de restrição tanto para a generalização, quanto para a especialização. Uma generalização pode ser total ou parcial. Ela é total quando a união de todas as instâncias das subclasses equivale ao conjunto completo de instâncias da superclasse. Neste caso, a notação gráfica é um ponto colocado no topo do triângulo. Os outros dois elementos de restrição tanto são o disjuncto e o sobreposto. Na notação gráfica para disjuncto é utilizado um triângulo em branco e, para sobreposto o triângulo é preenchido (BORGES *et al.*, 2005). A combinação de disjunção e totalidade gera os quatro tipos de restrição, as quais podem ser observadas na figura 12.

O penúltimo relacionamento é a agregação, uma forma de associação na qual se considera que um dos objetos é montado a partir de outros objetos. Ela pode ocorrer entre classes convencionais, georreferenciadas ou entre ambas. Na agregação entre classes georreferenciadas, utiliza-se a agregação espacial, um caso especial de agregação na qual, relacionamentos topológicos “todo-parte” são evidenciados. Isto quer dizer que a geometria de cada parte deve estar totalmente contida na geometria do todo, além disso, a superposição entre a geometria das

partes e a geometria do todo não é permitida (BORGES *et al.*, 2005). A figura 13 mostra o exemplo da classe quadra que é a agregação dos lotes.

FIGURA 13 – AGREGAÇÃO



FONTE: Davis Jr. & Laender (2000).

O último relacionamento é a generalização cartográfica. De acordo com Borges *et al.* (2005), a generalização cartográfica pode ser definida como “uma série de transformações que são realizadas sobre a representação da informação espacial, cujo objetivo é melhorar a legibilidade e aumentar a facilidade de compreensão dos dados por parte do usuário do mapa”. A generalização é utilizada para facilitar a compreensão dos dados e permitir a especificação de relacionamentos independentes para cada alternativa de representação. Conforme explicado anteriormente um objeto do mundo real pode ter diferentes representações, conforme a escala de visualização. Uma cidade pode ser representada como um ponto em um mapa de pequena escala, e como um polígono em um mapa de grande escala (BORGES *et al.*, 2005).

Na primitiva da generalização cartográfica incluída no modelo OMT-G, a superclasse não necessita ter uma representação específica, pois poderá ser visualizada com diferentes formas, de acordo com o especificado nas subclasses. Estas formas são geometricamente distintas, e podem herdar os atributos alfanuméricos da superclasse ou possuir seus próprios atributos (DAVIS JR. & LAENDER, 2000).

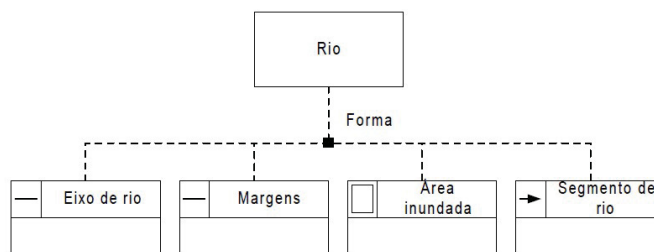
A generalização cartográfica pode acontecer em função da forma geométrica ou em função da escala. A variação em função da forma é utilizada para demonstrar a existência de múltiplas representações independente de escala para determinada classe. A variação em função da escala é utilizada para especificar diferentes aspectos geométricos de uma classe, e cada uma delas corresponde a uma faixa de escalas (BORGES *et al.*, 2005).

Conforme mostra a figura 14, a notação gráfica utilizada para a generalização cartográfica é um quadrado para ligar a superclasse a suas

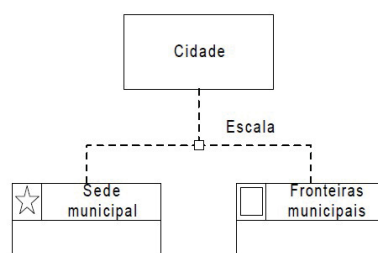


subclasses. O quadrado fica em branco quando as subclasses são disjuntas e, preenchido quando as subclasses são superpostas. Como se trata de um relacionamento espacial as linhas são pontilhadas e são utilizados os termos Escala e Forma para distinguir as variações da generalização (BORGES *et al.*, 2005).

FIGURA 14 – GENERALIZAÇÃO CARTOGRÁFICA NO MODELO OMT-G



(a) Variação de acordo com a forma (superposto)



(b) Variação de acordo com a escala (disjunto)

FONTE: Davis Jr. & Laender (2000).

Outra característica dos relacionamentos é a cardinalidade. A cardinalidade representa o número de instâncias de uma classe que podem estar associadas a uma instância de outra classe. São utilizados números para mostrar quantidades, reticências para expressar faixas e o asterisco para designar “muitos” ou “vários”. Por exemplo, um e muitos (1...\*), zero ou mais (0...\*), zero ou um (0...1), exatamente um (1) (BORGES *et al.*, 2005).

Conforme descrito nesta seção, os relacionamentos entre os objetos espaciais e não espaciais são desenhados, o que permite mais facilmente a percepção da interação entre eles.

Na próxima seção, foi apresentado o modelo conceitual referente aos dados do mapeamento colaborativo da plataforma *OpenStreetMap*, também utilizados nesta pesquisa.

## 2.3 MAPEAMENTO TOPOGRÁFICO EM GRANDES ESCALAS X MAPEAMENTO CADASTRAL NO BRASIL

O primeiro objetivo específico desta pesquisa foi efetuar um estudo semântico prévio para clarificar o entendimento do significado no uso dos termos mapeamento topográfico em grandes escalas e mapeamento cadastral no Brasil. No Brasil, estes dois conceitos são historicamente tratados como sinônimos. Esse estudo foi necessário para auxiliar no direcionamento do escopo da pesquisa e teve como resultado a geração de um artigo de revisão publicado no ano de 2019 na Revista Brasileira de Cartografia. O texto integral do artigo pode ser acessado e citado através do *link* <https://doi.org/10.14393/rbcv71n2-44528>.

É conveniente explicar que o objetivo inicial da pesquisa, seguindo a divisão da ET-EDGV 3.0 em Mapeamento Topográfico em Pequenas Escalas (MapTopoPE) e Mapeamento Topográfico em Grandes Escalas (MapTopoGE), era trabalhar com o MapTopoGE. Entretanto, durante o desenvolvimento da pesquisa entendeu-se que sob o para o aspecto da modelagem conceitual e semântico, as duas divisões do mapeamento não faziam sentido. Por isso, foram alinhadas as categorias das duas subdivisões do mapeamento.

## 2.4 MAPEAMENTO TOPOGRÁFICO OFICIAL NO BRASIL

No artigo citado na seção 2.3 foram diferenciados os conceitos de mapeamento topográfico em grandes escalas e de mapeamento cadastral. Nesta seção é introduzido o conceito de “oficial” à definição de mapeamento topográfico. Em seguida são apresentados os aspectos legais, as instituições responsáveis e as especificações empregadas na produção do mapeamento topográfico oficial em todas as escalas no país.

Os termos “mapeamento terrestre de referência”, “mapeamento sistemático terrestre”, “séries de cartas gerais” e “mapeamento topográfico terrestre” utilizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e pela Infraestrutura Nacional de Dados Espaciais (INDE), têm o mesmo significado de mapeamento topográfico, utilizado nessa tese.

#### 2.4.1 Conceito de mapeamento topográfico oficial

De acordo com (COLEMAN, 2013), o conceito de mapeamento topográfico oficial envolve algumas características que são semelhantes para a maioria dos países. As especificidades destas características no Brasil são tratadas nas próximas seções deste capítulo.

Primeiramente, o mapeamento topográfico oficial é produzido e administrado por instituições públicas ou privadas designadas para este propósito e com o objetivo de atender a um conjunto específico e definido de requisitos, que podem ser legais, administrativos ou econômicos. Além disso, os dados são obtidos por uma equipe de profissionais contratados para este fim. Deste modo, a produção do mapeamento topográfico oficial é baseada em métodos, padrões, especificações e técnicas pré-estabelecidas. A certificação da qualidade dos dados é estabelecida em diferentes níveis durante as etapas de produção, e quando são disponibilizados recebem metadados sobre esta característica (COLEMAN, 2013; GOODCHILD, 2009).

Dependendo do contexto, o mapeamento topográfico oficial pode ser disponibilizado gratuitamente, mas há casos nos quais existem custos e provavelmente restrições no acesso e no uso deste mapeamento. Os dados podem também ser protegidos de algum modo, por direitos autorais ou regidos por acordos formais ou licenças de uso. Além disso, parte do mapeamento oficial pode ter o acesso limitado por algumas instituições por razões de segurança, proteção dos dados ou benefícios econômicos da nação (COLEMAN, 2013; RAK, 2013).

Além destas características, o mapeamento topográfico oficial do Brasil, é financiado com recursos públicos, é direito de todos os cidadãos assegurado na Constituição Brasileira e na Lei de Acesso às Informações Públicas (BRASIL, 1988; BRASIL, 2011). O conhecimento do território, disponível e de fácil acesso a qualquer cidadão, é preceito fundamental para uma sociedade justa e deve ser entendido como parte essencial da construção da cidadania, tal como o acesso à educação, à saúde e ao trabalho (SLUTER, 2013).

## 2.4.2 Legislação, Instituições e Competências

De acordo com o artigo 2º, capítulo II, do decreto-lei nº 243, de 28 de Fevereiro de 1967, que fixa as Diretrizes e Bases da Cartografia Brasileira (BRASIL, 1967), as atividades cartográficas no Brasil são levadas a efeito através do Sistema Cartográfico Nacional (SCN), o qual é constituído pelas instituições nacionais, públicas e privadas, que têm por atribuição principal executar trabalhos cartográficos ou atividades correlatas. Entre estas instituições estão o Instituto Brasileiro de Geografia e Estatística (IBGE), a Diretoria do Serviço Geográfico do Exército (DSG) e a Comissão Nacional de Cartografia (CONCAR).

A CONCAR é incumbida de coordenar a execução da Política Cartográfica Nacional (BRASIL, 1967), e é constituída por representantes dos ministérios relacionados à defesa do país, da Associação Nacional de Empresas de Aerofotogrametria e do secretário-geral do Conselho Nacional de Geografia do IBGE. Entre as atribuições da CONCAR está a de elaborar as "Instruções Reguladoras das Normas Técnicas da Cartografia Terrestre Nacional", as quais foram estabelecidas através do decreto nº 89.817, de 20 de junho de 1984 (BRASIL, 1984). Embora a situação da CONCAR esteja indefinida atualmente (abril de 2020), a legislação não foi alterada.

Os trabalhos de natureza cartográfica realizados no território brasileiro obedecem às Normas Técnicas estabelecidas pelos órgãos federais competentes que cabem respectivamente (BRASIL, 1967, capítulo VIII, art. 15):

- “1. ao Conselho Nacional de Geografia, do Instituto Brasileiro de Geografia e Estatística, no que concerne à rede geodésica fundamental e às séries de cartas gerais, das escalas menores de 1:250.000;
- 2. à Diretoria do Serviço Geográfico, do Ministério da Guerra, no que concerne às séries de cartas gerais, das escalas de 1:250.000 e maiores;
- (...)

§ 2º As Normas Técnicas relativas às cartas temáticas e cartas especiais, não referidas neste artigo, são estabelecidas pelos órgãos públicos federais interessados, na esfera de suas atribuições, atendido o disposto no artigo 11.

§ 3º As Normas Técnicas de que trata o presente artigo serão publicadas pelos órgãos que as estabelecerem”.

No que se refere às escalas maiores do que 1:25.000 e, de acordo com o Plano de Ação da INDE (BRASIL, 2010a, p. 64; BRASIL, 2008) além da Diretoria do Serviço Geográfico a competência para a produção de dados geoespaciais do mapeamento terrestre de referência em grandes escalas cabe também aos “órgãos federais, estaduais e municipais através de contratação na iniciativa privada”.

A competência compartilhada traz consigo a questão da interoperabilidade para os dados. A adoção de normas e padrões viabiliza a interoperabilidade dos dados entre as instituições e evita a duplicidade e o desperdício de recursos na produção dos dados. Além disso, o mapeamento topográfico oficial é a base para a produção de outras informações, deste modo, os critérios adotados em relação a estes dados devem ser extremamente rigorosos, com o intuito de reduzir a propagação de erros aos produtos finais (BRASIL, 2010a).

Ademais a adoção de normas, padrões e especificações viabiliza, além da interoperabilidade, o estabelecimento de condições para controlar o processo de produção de dados espaciais e indica níveis de conformidade de qualidade dos produtos cartográficos concebidos (ARIZA-LÓPEZ, 2015; ARIZA-LÓPEZ, 2002).

#### 2.4.3 Normas, Padrões e Especificações

A Infraestrutura Nacional de Dados Espaciais (INDE) sumariza as normas, padrões e especificações necessárias à interoperabilidade e à qualidade dos dados produzidos pelas instituições de governo, de acordo com as competências definidas no decreto-lei nº 243 (BRASIL, 1967). Quanto ao mapeamento topográfico terrestre, o quadro 1 esclarece quais são eles (BRASIL, 2010a, p. 91).

A Especificação Técnica de Produtos de Conjuntos de Dados Geoespaciais (ET-PCDG) que tem como finalidade definir os padrões a serem observados na elaboração de produtos geoespaciais vetoriais e matriciais, encontra-se em sua 2ª edição (2016) (DSG, 2016c), aprovada pelo Exército.

A Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV) destina-se a padronizar estruturas de dados que viabilizem o compartilhamento, a interoperabilidade dos dados e a racionalização de recursos entre os produtores e usuários de dados e informações cartográficas. É o modelo

conceitual e semântico brasileiro o qual é tratado na próxima seção. A ET-EDGV encontra-se na seguinte situação:

- ET-EDGV versão 2.1.3 (2010), homologada pela CONCAR;
- ET-EDGV Força Terrestre, 2ª edição (2016), aprovada pelo Exército;
- ET-EDGV 3.0, 2ª edição (2017), homologada pela CONCAR (DSG, 2017).

A Especificação Técnica para a Aquisição de Dados Geoespaciais Vetoriais (ET-ADGV) define regras de aquisição da geometria dos dados de cada classe de objetos constante da ET-EDGV, bem como dos atributos essenciais à perfeita individualização das instâncias e os respectivos metadados, e encontra-se na seguinte situação:

- ET-ADGV versão 2.1.3 (2010), aprovada pelo Exército;
- ET-ADGV Defesa Força Terrestre, 1ª edição (2015), aprovada pelo Exército;
- ET-ADGV Defesa Força Terrestre, 2ª edição (2016), aprovada pelo Exército;
- ET-ADGV versão 3.0 (2018), em processo de homologação pela CONCAR (DSG, 2018).

## QUADRO 1 – MAPEAMENTO TERRESTRE: NORMAS, PADRÕES E ESPECIFICAÇÕES

**MAPEAMENTO TERRESTRE: NORMAS, PADRÕES E ESPECIFICAÇÕES**

Dados Geoespaciais		Especificação	Finalidade	Instituição Responsável	Situação/ Instituição Executora
<b>Cartografia Terrestre (Mapeamento Geográfico)</b>	Dados vetoriais, escalas menores que 1:250.000	Map. série Brasil (2ª versão Man. CIM, Doc. Téc. BCIM, e MD, v. 5.0)	Espec. Téc. que define um modelo de dados vetoriais para garantir a consistência lógica	IBGE Art. 8º Decreto Lei nº 243 (28/02/1967)	Em atualização IBGE
	Dados matriciais, escalas menores que 1:250.000	NI	NI	IBGE	NI
<b>Cartografia Terrestre (Mapeamento Topográfico)</b>	Dados matriciais, escalas 1:250.000 e maiores	ET-PCDG	Espec. Téc. que define os padrões dos produtos de conjuntos de dados geoespaciais do tipo carta matricial	Exército Brasileiro DSG Decreto Lei nº 243, (28/02/1967) Cap. IV, art. 6º, § 1º, letra b com o Cap. VII, art. 15, § 1º, nº 2	2ª edição aprovada pelo Exército (2016)
	Dados vetoriais, escalas 1:250.000 e maiores	ET-EDGV	Espec. Téc. que define um modelo conceitual e semântico para dados geoespaciais vetoriais		Versão 3.0 1ª edição homologada CONCAR (2017)
		ET-ADGV	Espec. Téc. que define regras de aquisição da geometria dos dados geoespaciais vetoriais		Versão 3.0 1ª edição aprovada pelo Exército (2018)
		ET-PCDG	Espec. Téc. que define os padrões dos produtos de conjuntos de dados geoespaciais vetoriais		2ª edição aprovada pelo Exército (2016)
		ET-RDG	Espec. Téc. que garante a consistência na representação das mesmas classes de objetos		Em elaboração (norma atual Manual T-34-700)
		ET-CQDG	Espec. Téc. que define os procedimentos para o controle de qualidade dos conjuntos de dados geoespaciais vetoriais		1ª edição aprovada pelo Exército (2016)
	Dados matriciais, escalas 1:10.000 e maiores		Espec. Téc. que define os padrões dos produtos de conjuntos de dados geoespaciais do tipo planta cadastral matricial	CONCAR	Em elaboração
<b>Cartografia Terrestre (Mapeamento Cadastral)</b>	Dados vetoriais, escalas 1:10.000 e maiores		Espec. Téc. que define os padrões das plantas cadastrais vetoriais	CONCAR	

NI = não identificado

FONTE: Adaptado de Brasil (2010).



A Especificação para a Representação de Dados Geoespaciais (ET-RDG) é um componente necessário para a representação gráfica de informações geoespaciais e encontra-se em elaboração pela Diretoria do Serviço Geográfico do Exército Brasileiro (DSG). Enquanto essa especificação não é lançada, a DSG recomenda a utilização do Manual Técnico T34-700 partes 1 e 2 (2ª edição) (DSG, 1998; DSG, 2000), onde estão definidos os símbolos clássicos do mapeamento sistemático e a ET-EDGV, onde estão definidas as classes a representar na modelagem conceitual brasileira. Para cada produto definido na ET-PCDG que requer representação gráfica, são definidos os estilos de representação gráfica de cada classe representada naquele produto.

A Especificação Técnica para o Controle de Qualidade dos Produtos de Conjuntos de Dados Geoespaciais (ET-CQDG) define os procedimentos para o controle de qualidade dos produtos da cartografia terrestre previstos na ET-PCDG e encontra-se em sua 1ª edição (DSG, 2016b), aprovada pelo Exército.

As especificações técnicas que definem os padrões dos produtos de conjuntos de dados geoespaciais do tipo planta cadastral matricial e vetorial ainda não foram elaboradas pela CONCAR.

Para o contexto dessa pesquisa é relevante a Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV). A ET-EDGV é o modelo conceitual brasileiro, o qual foi utilizado no alinhamento semântico proposto nesta pesquisa.

#### 2.4.3.1 Modelo Conceitual

Um modelo conceitual fornece uma descrição coerente e sistemática do conteúdo e organização de um conjunto de dados. Esse *schema* pode incluir a definição semântica das feições, o detalhamento dos nomes e tipos de atributos, um dicionário que define a estrutura dos objetos geográficos e seus atributos espaciais e não-espaciais, metadados proporcionando uma visão sinóptica das informações e um dicionário de termos relacionados (SONDHEIM *et al.*, 2015).

O modelo conceitual e semântico do mapeamento topográfico no Brasil, que é baseado no modelo OMT-G, é documentado pela Especificação Técnica para a Estruturação dos Dados Geoespaciais Vetoriais (ET-EDGV 3.0). Esta é a versão

mais recente da especificação, elaborada pelo Comitê Especializado para a Estruturação da Mapoteca Nacional Digital (CEMND) da CONCAR e lançada em dezembro de 2017.

Do ponto de vista da modelagem conceitual de dados geoespaciais, assunto introduzido na seção 2.2, a atual versão da ET-EDGV (3.0) tem enfoque na definição da estrutura de dados vetoriais, e visa atender às necessidades de produtores, desenvolvedores de SIG e usuários finais de dados geoespaciais (DSG, 2017).

Com base na OMT-G, o modelo conceitual brasileiro comporta classes georreferenciadas de geo-objetos com geometria dos tipos: Classe Georreferenciada Genérica; Classe Georreferenciada com Primitiva Geométrica Ponto; Classe Georreferenciada com Primitiva Geométrica Linha; e Classe Georreferenciada com Primitiva Geométrica Polígono. Quanto às classes georreferenciadas de geo-objetos com geometria e topologia: Classe Georreferenciada com Geometria e Topologia do tipo Nó; Classe Georreferenciada com Geometria e Topologia do tipo Linha Unidirecionada; e Georreferenciada com Geometria e Topologia do tipo Linha Bidirecionada. Já as classes georreferenciadas de geo-campos utilizadas na modelagem são: Classe Georreferenciada do tipo Isolinhas e Classe Georreferenciada do tipo Polígonos Adjacentes (DSG, 2017).

Os diagramas de classes são documentados e descrevem a estrutura e as Relações de Classes de Objetos (RCO) correspondentes, com seus atributos organizados por categorias de informação e listas de domínios dos atributos das classes de objetos utilizadas (DSG, 2017).

As classes de objetos na ET-EDGV 3.0 podem ser de dois tipos: georreferenciada ou convencional. Ambas são compostas por atributos e comportamentos. Os atributos descrevem as características estáticas dos objetos. Os comportamentos não são descritos e segundo a especificação esse assunto é tratado na ET-ADGV e na futura ET-RDG (DSG, 2017).

As chamadas classes convencionais não possuem propriedades geométricas. Essas classes estão relacionadas aos objetos espaciais e descrevem um conjunto de objetos com propriedades, comportamentos, relacionamentos e semântica semelhantes às das classes georreferenciadas. As classes georreferenciadas por sua vez, possuem propriedades geométricas e estão associadas ao mundo real (DSG, 2017).

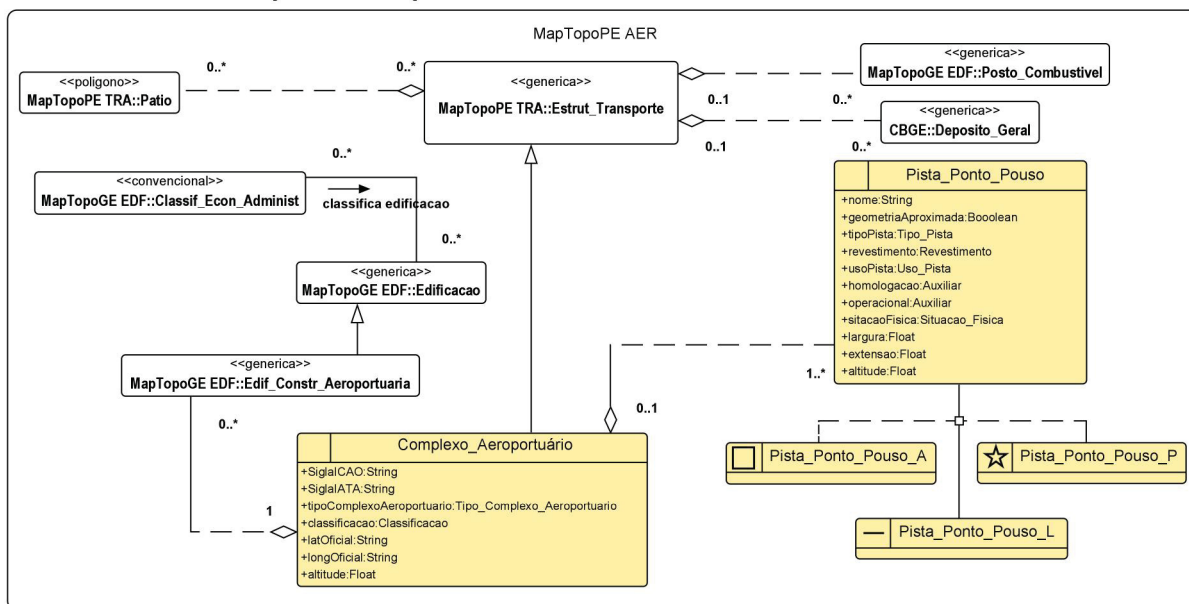
Entre as classes georreferenciadas, podem ocorrer as classes genéricas e complexas. A classe genérica ocorre quando algumas instâncias de classes georreferenciadas podem ser representadas por mais de uma primitiva geométrica, as quais ocorrem normalmente como superclasse em processos de generalização espacial ou para classes com geometria do tipo complexo. Uma classe com geometria do tipo complexo é aquela cuja geometria poderá ser constituída por mais de uma primitiva geométrica, isto poderá ocorrer em classes de objetos onde: pelo menos uma instância possua mais que uma primitiva geométrica e; as instâncias sejam representadas pela agregação de instâncias de classes de objetos com diferentes primitivas geométricas (DSG, 2017).

Cada classe de elementos contém atributos geométricos e semânticos associados que permitem a identificação e a classificação de cada elemento do mundo real representado no mapeamento (DSG, 2017).

A figura 15 mostra o diagrama de classes da categoria Sistema de Transporte Aeroportuário. Nela é possível observar que a classe *Complexo\_Aeroportuário* (uma classe complexa), agrega a classe *Pista\_Ponto\_Pouso*, que pode ter as três geometrias, dependendo da escala, e a classe *EDF\_Edif\_Constr\_Aeroportuaria* que é uma generalização da classe *Edificação*. A classe *Complexo\_Aeroportuário* também é uma generalização da classe *Estrut\_Transporte*, que por sua vez agrega as classes *TRA\_Estrut\_Transporte*, *TRA\_Patio*, *EDF\_Posto\_Combustivel* e *CBGE\_Deposito\_Geral*. É importante notar como as classes do Mapeamento Topográfico em Pequenas Escalas (MapTopoPE) e do Mapeamento Topográfico em Grandes Escalas (MapTopoGE) estão relacionadas entre si.

FIGURA 15 – DIAGRAMA DE CLASSES NA ET-EDGV 3.0

## 1.9 Sistema de Transporte/ Aeroportuário



- Classe de objetos original pertencente àquela categoria.
- Classe de objetos de outras categorias, necessárias a formação da categoria de informações.
- Classe de objetos abstrata, portanto, não instanciável.

FONTE: DSG (2017).

A figura 16 mostra a tabela da classe Complexo\_Aeroportuário, na qual é possível observar na parte superior, o nome da classe, a descrição semântica da feição, o código e a geometria prevista, complexa nesse caso. Logo abaixo, têm-se os atributos, o tipo (tamanho), a descrição do atributo, o domínio e os requisitos. Os atributos podem ser do tipo alfanumérico, real ou com geometria. Os alfanuméricos podem ser do tipo texto, numéricos ou associados a listas de domínio. O domínio se refere aos valores que podem ser assumidos em cada atributo. No caso do atributo tipo complexo aeroportuário, na seção 3.77 da ET-EDGV 3.0 são encontradas as três opções possíveis: aeroporto, aeródromo e heliporto com as respectivas descrições semânticas.

FIGURA 16 – TABELA DA CLASSE COMPLEXO AEROPORTUÁRIO

## 1.9.1 Complexo\_Aeroportuario

Classe	Descrição		Código	Geometria
Complexo_Aeroportuario	Complexo aeroportuário é um conjunto de elementos físicos cuja finalidade é apoiar as atividades relacionadas ao sistema aeroportuário, devendo o mesmo ser homologado pela ANAC.		1.9.1	C
Atributo	Tipo (tamanho)	Descrição	Domínio	Requisito
siglaCAO	Alfanumérico (04)	Indica o código do complexo aeroportuário estabelecido pela ICAO e constante no ROTAER.	A ser preenchido	0..1
siglaATA	Alfanumérico (03)	Indica a sigla oficial do complexo aeroportuário estabelecido pela IATA.	A ser preenchido	0..1
tipoComplexoAeroportuario	Tipo_Complexo_Aeroportuario	Identifica o tipo de complexo aeroportuário.	Seção 3.77	1
classificacao	Classificacao	Indica a classificação do complexo aeroportuário, quanto à operação.	Seção 3.9	1
latOficial	Alfanumérico (15)	Indica a latitude oficial (ROTAER) do ponto de referência do complexo.	A ser preenchido (-GGG°MM'SS.ssss") Ex.: -017°30'02.0059"	0..1
longOficial	Alfanumérico (15)	Indica a longitude oficial (ROTAER) do ponto de referência do complexo.	A ser preenchido (-GGG°MM'SS.ssss") Ex.: -048°30'02.0059"	0..1
altitude	Real	Indica a altitude oficial (ROTAER) do ponto de referência do complexo, em metros (m).	A ser preenchido Ex.: 1200.	0..1

Atributos herdados:

**Estrut\_Transporte:** nome, modalUso:Modal\_Uso="Aeroportuário", administracao, jurisdicao, concessionaria, operacional, situacaoFisica.

FONTE: DSG (2017).

As categorias de informação na ET-EDGV 3.0 são divididas em dois grupos: Mapeamento Topográfico em Pequenas Escalas (MapTopoPE) no qual as classes de objetos são agrupadas de modo a observar o aspecto funcional comum e; Mapeamento Topográfico em Grandes Escalas (MapTopoGE) que se refere a aquisição de dados em grande escala, geralmente planimétrica, com maior nível de detalhamento e alta precisão geométrica, especialmente usada na representação de feições das cidades e regiões metropolitanas, nas quais ocorre alta densidade de edificações e arruamentos. Esta versão da especificação tem 19 categorias de informação, sendo 14 categorias no grupo das escalas pequenas e 5 categorias no grupo das escalas grandes. Segundo a ET-EDGV 3.0 (DSG, 2017) a modelagem das categorias de informação utiliza os dois esquemas MapTopoPE e MapTopoGE, “mantendo a coerência das informações geoespaciais entre as escalas, aplicando o conceito de generalização espacial” (DSG, 2017).

Entretanto, esta característica da ET-EDGV 3.0 de dividir as categorias de informação em Mapeamento Topográfico em Pequenas Escalas (MapTopoPE) e Mapeamento Topográfico em Grandes Escalas (MapTopoGE), comporta uma incoerência conceitual, dado que a representação das entidades no banco de dados geoespacial é independente da escala. O que muda é a sua apresentação de acordo com as faixas de escalas, conforme explicado na seção 2.2.1 sobre os níveis de abstração de aplicações geoespaciais. A variação de acordo com a forma é

utilizada para registrar a existência de múltiplas apresentações para uma classe, independente de escala (DAVIS & LAENDER, 2000).

O histórico de desenvolvimento das versões da ET-EDGV mostra como esta incoerência conceitual foi introduzida.

O início dos esforços para elaboração da primeira versão da especificação data de junho de 1997. Na época a Sub-Comissão de Normas da CONCAR criou o Comitê Especializado para Estudo do Padrão de Intercâmbio de Dados Cartográficos Digitais (CEPAD) para estabelecer um padrão para orientar o intercâmbio de dados cartográficos digitais entre as instituições produtoras de dados geoespaciais governamentais. O trabalho da DSG e do IBGE para criação da especificação tiveram seu primeiro resultado em 2005, tendo a sua primeira versão homologada pela CONCAR em 2006 para uso em caráter provisório. A segunda versão, resultante da discussão e aperfeiçoamento da versão de 2005 pelos órgãos do Sistema Cartográfico Nacional (SCN), submetida à consulta pública e homologada pela CONCAR em 2010, foi a ET-EDGV 2.1.3 (DSG, 2010). Nesta primeira versão, as 13 categorias de informação não foram divididas em pequenas e grandes escalas. De acordo com a especificação (DSG, 2010), a modelagem conceitual e lógica da estrutura dos dados geoespaciais são compatíveis com as escalas de 1:25.000 e menores.

A versão seguinte da especificação, a ET-EDGV de Defesa da Força Terrestre (ET-EDGV-DefesaFT – 1ª Parte, 1ª edição), segundo a DSG (DSG, 2015) é uma extensão da ET-EDGV 2.1.3, à qual foi agregada a modelagem conceitual e lógica para os dados do Mapeamento Topográfico em Grandes Escalas (MapTopoGE) e da geoinformação temática pertinente a Defesa e a Segurança. A ET-EDGV-DefesaFT foi desenvolvida para o contexto das ações de segurança realizadas pelo Exército Brasileiro nos eventos de grande porte sediados no país em 2013 (Copa das Confederações), 2014 (Copa do Mundo) e 2016 (Jogos Olímpicos).

O planejamento e a execução de operações nestes eventos demandam o uso de geoinformação temática de defesa e segurança, além de uma base de dados em grandes escalas, caso das bases cartográficas produzidas pelas prefeituras municipais. Todavia, as bases de dados do Mapeamento Topográfico em Grandes Escalas de cada município, fornecidas ao Exército pelas prefeituras, não seguiam um padrão único, o que impossibilitava a interoperabilidade dos sistemas de defesa e segurança, bem como a sua integração em um mesmo banco de dados

geoespaciais, razão pela qual a DSG, elaborou a ET-EDGV Defesa F Ter. Esta versão da especificação tem 24 categorias de informação, as quais foram então divididas nos grupos de (MapTopoPE) (16 categorias), (MapTopoGE) (8 categorias) e mais 19 categorias de dados temáticos de defesa, as quais são de uso exclusivo do Exército Brasileiro e que são apenas citadas. A partir desta versão a modelagem conceitual e lógica da estrutura dos dados geoespaciais são relativas às escalas de 1:1.000 e menores.

A ET-EDGV F Ter 2ª edição (DSG, 2016a), foi lançada no primeiro semestre de 2016. Nela foram agregadas sugestões da CONCAR e aspectos relacionados as experiências acumuladas durante o período de uso da ET EDGV Defesa F Ter 1ª Edição pelo Exército Brasileiro. Esta versão da especificação tem 19 categorias de informação, as quais foram então divididas nos grupos de (MapTopoPE) (14 categorias), (MapTopoGE) (5 categorias) e mais 25 categorias de dados temáticos de defesa.

No segundo semestre de 2014, a CONCAR passou a utilizar a ET-EDGV-DefesaF Ter como referência para elaboração da futura versão 3.0, abrangendo o Mapeamento Topográfico em Grandes Escalas (DSG, 2015).

Deste modo, foi a partir daí que a incoerência conceitual foi introduzida, pois a estruturação e a modelagem agregaram uma divisão por grupos de escalas e não apenas as novas categorias, já que a representação das entidades no banco de dados geoespacial é independente da escala. Um aspecto que demonstra que esta forma de representação não é a ideal é que foi criada uma categoria, dentro do grupo das grandes escalas, chamada de “Classes Base do Mapeamento Topográfico em Grandes Escalas” a qual repete classes que fazem parte do grupo das pequenas escalas. Possivelmente, se a estruturação tivesse sido feita prevendo generalização no mapeamento e em diversas escalas, esta divisão não aconteceria.

Sob o aspecto semântico, embora existam problemas pontuais, a ET-EDGV 3.0 fornece as descrições dos significados dos conceitos e feições de modo organizado permitindo a estruturação e o compartilhamento do conhecimento nela contido a respeito do mapeamento topográfico oficial brasileiro.

Paralelamente ao lançamento das edições da ET-EDGV foi também disponibilizado o *DSGTools*, um *plugin* para o *software* livre de SIG, *QGIS*, gratuito e aberto, desenvolvido em linguagem *Python* pela DSG. O *DSGTools* fornece um conjunto de funcionalidades destinadas à criação e à manutenção de bancos de



dados geoespaciais, tanto usando o SGBD *PostGIS* quanto o SGBD *SpatiaLite*, de acordo com as versões da especificação. Esse *plugin* permite a importação de conjuntos de dados e a sua padronização de acordo com a ET-EDGV.

Na próxima seção foi abordada a questão da qualidade no mapeamento topográfico oficial e as especificações nacionais pertinentes a esse tema.

#### 2.4.3.2 Qualidade

A questão da qualidade é tratada nessa seção com o intuito de apresentar os objetivos e os mecanismos de avaliação no contexto do mapeamento topográfico oficial. Nessa tese não houve o objetivo de fazer a avaliação de qualidade do mapeamento colaborativo em comparação ao mapeamento oficial. Contudo, como existem muitos questionamentos em relação à qualidade do mapeamento colaborativo (os quais serão tratados na seção 2.5.2.2 Qualidade no capítulo referente ao mapeamento colaborativo do *OSM*), essa seção consta da tese como parâmetro para demonstrar a diferença dos mecanismos de avaliação entre os dois tipos de mapeamento.

Dados geoespaciais são o resultado de um processo de produção e, o modo como este ocorre, afeta a qualidade dos dados finais. Historicamente, a produção de dados geoespaciais de referência sempre foi de responsabilidade das agências nacionais de mapeamento como o *US Geological Survey* (USGS) nos Estados Unidos, o *Ordnance Survey* (OS) na Grã-Bretanha e o Instituto Brasileiro de Geografia e Estatística (IBGE) no Brasil. Cada uma delas utiliza determinada especificação para garantir a qualidade dos dados produzidos (COLEMAN, 2013).

Contudo, nas últimas décadas o uso de Sistemas de Informações Geográficas como ferramenta de tomada de decisões e planejamento pelos mais diversos setores e usuários, a implantação de Infraestruturas de Dados Espaciais (IDEs) em diferentes níveis hierárquicos, bem como a produção e a utilização de dados geoespaciais por usuários não-especialistas fizeram com que aumentasse a preocupação com a qualidade dos dados geoespaciais (UREÑA-CÁMARA *et al.*, 2018; JOKSIC & BAJAT, 2004).

Até então, a qualidade dos dados bem como sua avaliação eram de responsabilidade dos produtores oficiais. Neste novo cenário a avaliação da



qualidade dos dados também é feita pelo usuário, pois é este quem analisa se a qualidade dos dados atende ao uso por ele requerido. A melhor qualidade neste caso, pode não ser a *performance* máxima, mas aquela especificada para responder às necessidades do usuário, ou seja, uma informação considerada imprecisa para um usuário pode ser suficiente para outro (UREÑA-CÁMARA *et al.*, 2018; WEBER, *et al.*, 1999).

A preocupação com a qualidade, bem como sua avaliação no cenário atual, pode ser demonstrada pelas recentes atualizações das especificações relativas à qualidade de dados geoespaciais, como é o caso da substituição das normas ISO 19113:2002 *Geographic information – Quality principles* (ISO/TC211, 2002), ISO 19114:2003 *Geographic information – Quality evaluation procedures* (ISO/TC211, 2003) e ISO/TS 19138:2006 *Geographic information – Data quality measures* (ISO/TC211, 2006) pela ISO 19157:2013 *Geographic information – Data quality* (ISO/TC211, 2013). Esta última, de acordo com a sua especificação, comporta a perspectiva do usuário, pois (tradução livre):

"Se aplica aos produtores de dados fornecendo informações de qualidade para descrever e avaliar o quão bem um conjunto de dados está em conformidade com a especificação do produto e para usuários de dados procurando determinar se dados geoespaciais específicos têm qualidade suficiente para sua aplicação particular". (ISO/TC211, 2013).

Neste contexto, o papel do produtor de dados geoespaciais mudou: mais do que produzir dados oficiais, sua preocupação passou a ser documentar a qualidade dos dados com vistas à sua adequada utilização ou "*truth-in-labelling*". O chamado "*truth-in-labelling paradigm*" (verdade no paradigma de rotulagem) entende o erro como inevitável e molda o problema da qualidade dos dados em termos do uso indevido decorrente do conhecimento incompleto das limitações dos dados (LACASTA *et al.*, 2014; VEREGIN, 1999).

Deste modo, o conceito de qualidade se refere então à conformidade dos dados com uma determinada especificação e, além disso, à sua capacidade de atender ao uso requerido pelos usuários.

No Brasil, a qualidade de dados geoespaciais é abordada na Especificação Técnica para Controle de Qualidade de Dados Geoespaciais (ET-CQDG) (DSG, 2016b), a qual está em sua primeira versão e foi lançada em 2016. A ET-CQDG foi

elaborada em conformidade com a norma ISO 19157:2013 *Geographic information – Data quality* e está em sintonia com as demais normas nacionais.

A ET-CQDG tem como objetivo fornecer uma forma padronizada para avaliar a qualidade dos produtos de conjuntos de dados geoespaciais integrantes do Sistema Cartográfico Nacional (SCN). Ela estabelece definições relativas à avaliação da qualidade de produtos geoespaciais; descreve as medidas e os procedimentos que devem ser usados para avaliar a qualidade; apresenta os parâmetros de conformidade para os produtos de referência no país; e define como deve ser reportada a qualidade (DSG, 2016b).

A especificação abrange os cinco elementos da qualidade (completude, consistência lógica, acurácia posicional, acurácia temporal, acurácia temática), os quais são descritos a seguir, e apresenta os procedimentos de avaliação da qualidade para os tipos de produtos previstos Especificação Técnica para Produtos de Conjuntos de Dados Geoespaciais (ET-PCDG).

A **completude** se refere à presença ou ausência de feições no conjunto de dados. As feições que devem estar presentes em cada produto dependem da escala e seguem uma especificação de aquisição de dados. Se alguma feição que deveria estar representada, não foi representada, ocorre o erro de omissão, o qual deprecia a qualidade do produto. Se alguma feição que não deveria estar representada, foi representada, ocorre o erro de comissão, o qual, do mesmo modo deprecia a qualidade do produto (ISO/TC211, 2013; DSG, 2016b).

A **consistência lógica** se refere ao cumprimento das regras lógicas da estrutura dos dados, dos atributos e das suas relações. As regras lógicas seguem uma especificação de estruturação de dados nos níveis conceitual, físico e lógico. A consistência lógica se subdivide em quatro elementos: o primeiro é a consistência conceitual que diz respeito à aderência ao modelo conceitual estabelecido para os dados; o segundo é a consistência de domínio segundo a qual, os campos nas tabelas de atributos devem ser preenchidos de acordo com os valores estabelecidos para aquele tipo de informação; o terceiro é a consistência de formato, que diz respeito ao grau com que os dados são armazenados de acordo com a estrutura física do conjunto de dados; o quarto é a consistência topológica que se refere à correção das características topológicas estabelecidas para o conjunto de dados (ISO/TC211, 2013; DSG, 2016b).

A **acurácia posicional** se refere à qualidade da posição geoespacial das coordenadas do conjunto de dados. É classificada em: absoluta ou externa, quando ocorre discrepância entre as coordenadas da representação da feição e as coordenadas das feições na fonte de maior precisão; relativa ou interna, quando ocorre discrepância entre as posições relativas dos objetos e às posições relativas reais ou aceitas como sendo certas e; do *grid*, que consiste da aplicação do conceito de acurácia posicional absoluta para dados em malha regular, como nos Modelos Digitais de Elevação (MDE) (ISO/TC211, 2013; DSG, 2016b).

A **acurácia temporal** se refere à exatidão da dimensão temporal dos dados geoespaciais. É classificada em: exatidão de uma medida de tempo, que se refere a acurácia de um determinado atributo temporal (evento ou período) em relação ao valor real; consistência temporal, verifica se a ordem em que ocorreram os eventos/períodos está coerente com o tempo e; validade temporal, que diz respeito à validade de uma medida em relação ao tempo (ISO/TC211, 2013; DSG, 2016b).

A **acurácia temática** se refere à correta interpretação das feições e atributos, bem como à sua inclusão nas classes contidas no modelo conceitual. É classificada em: acurácia da classificação, obtida pela comparação das classes e/ou atributos encontrados no conjunto de dados geoespaciais e o modelo de dados; acurácia de atributos não quantitativos, que é obtida pela avaliação dos atributos que não podem ser contados, quando comparados com os atributos das mesmas feições na fonte de maior precisão e; acurácia de atributos quantitativos, que é obtida pela avaliação dos atributos que podem ser contados, quando comparados com os atributos das mesmas feições na fonte de maior precisão (ISO/TC211, 2013; DSG, 2016b).

De acordo com a ISO 19157:2013 (ISO/TC211, 2013), cada elemento da qualidade possui uma medida, um método de avaliação e um ou mais resultados. Estes resultados devem ser comparáveis com outros produtos do mesmo tipo e seus valores representam o grau de aderência às regras estabelecidas pela especificação. Podem ser avaliados: uma lista de feições, os atributos das mesmas ou atributos das feições e seus relacionamentos, uma parte de um conjunto de dados, ou o conjunto como um todo. Os itens a serem avaliados recebem a denominação de escopo.

O processo de avaliação da qualidade de dados geoespaciais consiste em (ISO/TC211, 2013):

- Definir a unidade de qualidade (escopo + elemento);
- Especificar a medida de qualidade e seus parâmetros;
- Escolher um tipo de avaliação: direto interno, direto externo, ou indireto; e
- Escolher um método de avaliação: inspeção completa, inspeção por amostragem, agregação/derivação, ou avaliação indireta;
- Determinar a saída do procedimento de avaliação (tabela, relatório, etc.) levando em consideração o nível de conformidade do produto.

As medidas de avaliação utilizadas para quantificar os elementos de qualidade dependem de cada produto a ser avaliado e derivam de medidas básicas como: indicador de erro (*booleano*), indicador de acerto (*booleano*), contagem de erro (inteiro), taxa de erro (real), taxa de acerto (real).

Os tipos e os métodos de avaliação dependem dos tipos de conjuntos de dados (como conjunto de dados geoespaciais vetoriais, carta topográfica, ortomagem, etc.) e da respectiva escala destes dados.

Para que os usuários possam acessar a qualidade dos dados através dos cinco elementos que a compõem, e avaliar se estes dados têm os requisitos necessários ao uso por ele pretendido é necessário que estas informações sejam documentadas através de metadados. As informações descritas pelos metadados possibilitam avaliar com que grau um conjunto de dados satisfaz as necessidades de uma aplicação e inferir o produto resultante de uma análise efetuada com eles. O padrão de metadados nacional é o Perfil de Metadados Geoespaciais do Brasil (Perfil MGB), o qual será apresentado na seção 2.4.3.3.

Outro documento recentemente lançado no país que trata da qualidade é o Manual Técnico de Avaliação da Qualidade de Dados Geoespaciais do IBGE (IBGE, 2017). O manual destaca que a avaliação de qualidade pode ser também aplicada ao processo de produção e não apenas ao produto final. Além disso, defende que (IBGE, 2017) “a ideia de controlar um processo para conseguir melhorias é totalmente diferente daquela de inspecionar produtos para identificar os não conformes, embora os dois procedimentos utilizem em parte as mesmas ferramentas estatísticas”.

De acordo com IBGE (2017), a avaliação de qualidade de um produto final, tem como objetivo determinar estatística e objetivamente um resultado final, sem se preocupar com o processo que o gerou. Já o controle de qualidade de um processo

tem como objetivo a avaliação do processo de produção como fornecimento de um bem ou serviço, e consequentemente da qualidade da organização produtora. Nessa linha de pensamento, se o processo de produção for correto e adequado, o resultado também o será. A avaliação da qualidade do processo de produção inclui então, aspectos importantes do ponto de vista do usuário, como prazo de entrega, garantia, serviço pós-venda, atualização do produto, preço, instruções de uso, entre outros. Contudo, no documento é abordada apenas a inspeção de qualidade para produtos.

O Manual incorpora a usabilidade, elemento de qualidade não abordado na ET-CQDG. A avaliação da usabilidade é baseada nos requisitos específicos dos usuários, nos quais todos os elementos de qualidade de dados espaciais podem ser usados. A usabilidade avalia se um produto atende a especificações dos usuários, por meio de indicadores como eficácia, eficiência e satisfação em um contexto de uso específico (IBGE, 2017).

Deste modo, o Manual Técnico de Avaliação da Qualidade de Dados Geoespaciais do IBGE pode ser usado como um documento complementar da ET-CQDG.

Na próxima seção será tratado o Perfil de Metadados Geoespaciais do Brasil (Perfil MGB), importante na forma como deve ser reportada a qualidade. No relatório de qualidade, os resultados podem ser informados como metadados ou como um relatório independente. No caso da utilização de metadados a qualidade é informada utilizando-se o Perfil MGB.

#### 2.4.3.3 Metadados

O Perfil de Metadados Geoespaciais do Brasil (Perfil MGB) (CONCAR, 2011) foi o primeiro padrão lançado no contexto da Infraestrutura Nacional de Dados Espaciais (INDE). Entre os diversos componentes de uma infraestrutura desta natureza, os metadados são os elementos centrais na dinâmica de disseminação e uso efetivo dos dados geoespaciais. O decreto de implantação da INDE define os metadados como (BRASIL, 2008, Art. 2º):

“Um conjunto de informações descritivas sobre os dados, incluindo as características de seu levantamento, produção, qualidade e estrutura

de armazenamento, essenciais para promover a sua documentação, integração e disponibilização, bem como possibilitar sua busca e exploração”.

O Perfil MGB, foi homologado pela Resolução nº 1 de 30/11/2009 (BRASIL, 2009) e inclui a maioria das seções de metadados presentes na norma ISO 19115:2003 *Geographic Information – Metadata*. O perfil brasileiro, a exemplo de outros padrões, está conceituado e estruturado em seções com funções específicas de: identificar o produtor e a responsabilidade técnica de produção; padronizar a terminologia utilizada; garantir o compartilhamento e a transferência de dados; viabilizar a integração de informações; possibilitar o controle de qualidade; e garantir os requisitos mínimos de disponibilização.

O Perfil MGB utiliza a modelagem da *Unified Modeling Language* (UML) para representar suas seções, entidades e elementos de metadados. Ele contempla as seguintes seções de metadados geoespaciais:

- Informações do Conjunto de Entidades de Metadados: define metadados de um produto;
- Informações de Identificação: informação para identificar univocamente um produto;
- Informações de Restrições: restrições legais e de segurança no acesso e no uso dos dados;
- Informações de Qualidade dos Dados: descreve sua linhagem e qualidade. O anexo 5.4 do Perfil MGB contém as orientações para preenchimento da linhagem e relatórios de qualidade (para as grandes etapas dos principais processos de produção);
- Informações de Manutenção dos Dados: descreve práticas de manutenção e atualização;
- Informações de Representação Espacial: descreve mecanismo usado para representação dos dados (matricial ou vetorial);
- Informações do Sistema de Referência: sistema de referência espacial e temporal usado;
- Informações de Conteúdo: descreve conteúdo do(s) catálogo(s) de abrangência e de feições usado(s) para definir feições de dados geoespaciais;
- Informações do Distribuidor: informações do distribuidor e métodos de acesso.

A CONCAR (CONCAR, 2011) recomenda que na descrição de dados geoespaciais de referência, caso do mapeamento topográfico, seja preenchido o Perfil MGB completo. Mas, para fomentar a cultura de documentação de quaisquer produtos de geoinformação através de metadados, também disponibiliza dentro do mesmo padrão o Perfil MGB Sumarizado. o Perfil MGB Sumarizado contém 23 elementos, alguns obrigatórios, outros condicionais e os demais opcionais.

No próximo capítulo foi tratado o mapeamento colaborativo e os aspectos do mapeamento do *OpenStreetMap* relevantes para o alinhamento semântico com o mapeamento topográfico oficial brasileiro proposto nessa pesquisa.

## 2.5 MAPEAMENTO COLABORATIVO

### 2.5.1 Conceito de mapeamento colaborativo

O termo *Volunteered Geographic Information (VGI)*, Informações Geográficas Voluntárias ou mapeamento colaborativo foi criado por Michael F. Goodchild em 2000, para designar o crescente interesse de pessoas sem conhecimento especializado, em criar voluntariamente informações geoespaciais na *Web*. De acordo com ele (GOODCHILD, 2007):

“sites como *Wikimapia* e *OpenStreetMap* estão "empoderando" os cidadãos a criarem um mosaico global de informações geográficas, enquanto o *Google Earth* e outros globos virtuais estão encorajando as pessoas a desenvolverem aplicativos utilizando seus próprios dados”.

As informações geográficas voluntárias evoluíram da chamada *Web 2.0*, no qual os usuários passaram de atores passivos, apenas visualizando conteúdos na *internet (Web 1.0)*, para atores ativos, a princípio preenchendo formulários e depois, criando *sites* com conteúdo inteiramente gerado por eles, como *blogs* e *wikis* (GOODCHILD, 2007).

Atribui-se essa evolução ao desenvolvimento das tecnologias como é o caso da criação e da popularização do GPS (*Global Positioning System*), que possibilitou o georreferenciamento de pontos sobre a superfície terrestre de modo bastante simples, permitindo inserir a localização de fotos e rotas à partir de câmeras e dispositivos móveis; o incremento na qualidade gráfica dos *hardware* para

visualização e; a comunicação através de banda larga por cabos, satélites, e telefonia, aumentando a velocidade de processamento nas tarefas no computador (GOODCHILD, 2007).

O fenômeno *VGI* preconizou os impactos que vêm ocorrendo nos sistemas de informação geográfica, na comunidade de informação geoespacial e na sociedade como um todo. O declínio na produção e atualização do mapeamento sistemático em muitos países, a emergência das IDEs e a adoção de padrões são meios de descentralizar a produção de dados geoespaciais e convidar outros membros da sociedade a colaborarem voluntariamente na construção da representação do espaço geográfico (OLTEANU-RAIMOND, *et al.*, 2017; GIRRES & TOUYA, 2010).

Contudo, é possível investigar mais profundamente as informações geográficas voluntárias refletindo sobre conceitos e fenômenos prévios, bem como as motivações que envolvem a geração de conhecimento de modo coletivo por pessoas com diferentes habilidades e interesses. Dentre estes conceitos prévios, foi aqui abordado o conceito de Inteligência Coletiva.

O termo Inteligência Coletiva (CI) foi cunhado por Surowiecki (2005) em seu livro *“Wisdom of Crowds”* ou sabedoria das multidões, no qual ele ilustra diversas situações nas quais um grupo de pessoas sem conhecimento especializado pode solucionar um problema específico melhor do que um especialista. Entre as situações ilustradas no livro está a de um cientista de 85 anos com um trabalho notório em estatística e hereditariedade que no ano de 1885 faz um experimento na Feira Internacional de Londres. Após ter havido uma competição com premiação para adivinhar qual seria o peso de um boi no qual 800 pessoas presentes fizeram suas apostas, ele comprou os tíquetes e realizou uma série de testes estatísticos com eles, calculando por fim a média dos palpites que resultou em 1.197 quilos, apenas um quilo a menos do que o verdadeiro peso do boi (SUROWIECKI, 2005).

De acordo com o autor Surowiecki (2005) (tradução livre): “sob as circunstâncias adequadas, grupos são notavelmente inteligentes, e são frequentemente mais inteligentes que as pessoas mais inteligentes entre eles”.

Para que as “multidões” tenham a sabedoria requerida para resolver problemas, são necessárias quatro condições: diversidade de opinião, pois cada pessoa deve ter alguma informação particular, mesmo que seja apenas uma interpretação excêntrica dos fatos; independência, pois a opinião das pessoas não



deve ser determinada pelas opiniões daqueles que estão ao seu redor; descentralização, pois as pessoas são capazes de se especializar e gerar conhecimento local; e agregação, porque existem mecanismos para transformar julgamentos particulares em decisões coletivas (SUROWIECKI, 2014).

Se um grupo satisfaz essas quatro condições, seu julgamento provavelmente será preciso. Por quê? No fundo, a resposta reside em uma obviedade matemática. Se pedirmos a um grupo suficientemente diversificado e independente de pessoas para fazer uma predição ou uma probabilidade e então fizermos a média dessas estimativas, os erros que cada um poderia cometer seriam anulados. O palpite de cada pessoa tem dois componentes: informação e erro. Subtraindo-se o erro, ficamos com a informação. É possível que o julgamento de um grupo seja ruim, mesmo que seus erros sejam anulados. Para que um grupo seja sábio, deve haver pelo menos alguma informação na equação informação menos erro (SUROWIECKI, 2005). Mais adiante, na seção 2.5.2.2 é tratada a qualidade no mapeamento colaborativo, as informações aqui apresentadas podem auxiliar na compreensão desta questão.

O mecanismo de busca do *Google* que utiliza o algoritmo *PageRank*, desenvolvido por seus fundadores Sergey Brin e Lawrence Page em 1995, é um exemplo de utilização da inteligência coletiva. Ele funciona da seguinte forma: o *Google* interpreta o *link* da página A para a página B como se fosse um voto da página A para a página B. Ele acessa a importância de uma página pelos votos que ela recebe, mas o *Google* olha para além do volume de votos ou *links*, ele também analisa a página que concede o voto. Votos concedidos por páginas que são elas mesmas “importantes” têm peso maior e ajudam a tornar outras páginas importantes. Esta é uma forma de utilizar o conhecimento coletivo gerado por milhões de pessoas fazendo *sites* para outros propósitos e, aproveitar este conhecimento para produzir respostas para as questões que digitamos no motor de busca (SUROWIECKI, 2005; MALONE, 2008; BRIN & PAGE, 2012).

A *Wikipedia* utiliza uma tecnologia menos sofisticada que o *Google*, mas em compensação faz uso de princípios organizacionais e técnicas motivacionais para conseguir que milhares de pessoas pelo mundo voluntariem seu tempo para criar uma coletânea de conhecimento *online* (MALONE, 2008).

Atualmente, um grande número de companhias como a *Hewlett Packard*, *Eli Lilly* e a *Google*, estão utilizando os chamados mercados preditivos ou mercados

especulativos – um método para capturar a inteligência coletiva – nos quais as pessoas vendem e compram previsões sobre eventos futuros (como as vendas dos seus produtos) e obtêm na maioria dos casos, resultados mais precisos que os mercados tradicionais de pesquisas, de previsões eleitorais ou outras técnicas (MALONE, 2008).

Surowiecki (2005) aborda três tipos de problemas, os quais ele denomina de problemas de cognição, problemas de coordenação e problemas de cooperação.

Os problemas de cognição são aqueles que têm ou terão soluções definitivas como, por exemplo, “quem vencerá o campeonato brasileiro de futebol este ano?”, ou, questões para as quais pode não haver uma única resposta correta, mas para as quais algumas respostas são melhores que outras, como “qual seria o melhor lugar para construir uma academia ao ar livre?”.

Os problemas de coordenação requerem membros de um grupo (mercado, estudantes procurando uma festa) para imaginar como coordenar o comportamento entre eles sabendo que todos estão tentando fazer a mesma coisa, por exemplo, “como compradores e vendedores se encontram e negociam um preço justo?” ou “como dirigir em segurança em tráfego intenso?”.

Os problemas de cooperação envolvem o desafio de manter a motivação individual e/ou fazer com que pessoas desconfiadas trabalhem juntas, mesmo quando o interesse próprio parece impor que não se deve tomar parte, como é o caso do pagamento de impostos, lidar com a poluição ou concordar na definição de quais contas a se pagar são justas.

### 2.5.2 *OpenStreetMap*

Entre as iniciativas de mapeamento colaborativo, a plataforma *OpenStreetMap* (OSM) foi escolhida porque é uma plataforma aberta e livre (característica explicada nos próximos parágrafos) e porque sua proposta é a de ser uma alternativa cujo foco principal é armazenar informações de referência, em lógica similar ao conceito de mapeamento topográfico tradicional. Além disso, o OSM conta com uma comunidade bastante numerosa de usuários, tem todos os seus componentes bem documentados e disponíveis na *internet* e têm sido amplamente

investigado e avaliado por inúmeras pesquisas científicas, aspectos que fornecem subsídios que permitem fundamentar sua utilização (OSM, 2020).

O *OpenStreetMap* (OSM) é um projeto de mapeamento colaborativo global, fundado em julho de 2004 por Steve Coast, um empreendedor britânico graduado em Ciência da Computação na *University College London* (UCL). O Reino Unido foi o foco inicial do mapeamento pelo OSM, já que o *Ordnance Survey* (OS), agência oficial de mapeamento da Grã-Bretanha, produzia conjuntos de dados geoespaciais, mas a exemplo de outras agências de mapeamento oficiais no mundo, não os distribuía gratuitamente (OSM, 2020).

Em 2006, foi estabelecida a *OpenStreetMap Foundation*, uma organização sem fins lucrativos para dar suporte financeiro e jurídico ao projeto. O *OpenStreetMap* é constituído por dados abertos, ou seja, qualquer pessoa tem a liberdade de usar os dados para qualquer fim, desde que credite a autoria aos contribuidores do OSM. Se os dados forem alterados ou, se for criado algo novo com eles, o produto resultante poderá ser distribuído apenas sob a mesma licença, ou seja, livres de barreiras à utilização através de direitos autorais. Os dados anteriores a setembro de 2012 são disponibilizados sob a licença *Creative Commons Attribution Share Alike* (CC-BY-AS 2.0) ou atribuição-compartilhamento pela mesma licença. Os dados posteriores a setembro de 2012 são disponibilizados sob a licença *Open Data Commons Open Database License* (ODbL 1.0) (OSM FOUNDATION, 2020).

Uma preocupação das agências de mapeamento que estudam agregar os dados do OSM aos dados oficiais ou disponibilizar os dados oficiais para o OSM são os conflitos de licenciamento. Todos os dados integrados com os dados do *OpenStreetMap* devem ser distribuídos usando a mesma licença ODbL 1.0 e, portanto, requerem a atribuição aos colaboradores do OSM. Para os usuários de mapas, o licenciamento não representa um problema desde que os dados sejam abertos, mas para uma agência de mapeamento é importante preservar os direitos autorais dos dados. Rönneberg *et al.*, (2019) dizem que uma agência de mapeamento pode oferecer seus dados topográficos como dados abertos sob as licenças CCBY e, nesse caso, eles não poderiam ser mesclados com os dados sob a licença ODbL.

O OSM é desenvolvido por uma comunidade voluntária que produz e mantém atualizados os dados geoespaciais de uma infinidade de feições em todo o

mundo (OSM, 2020). Até abril de 2020 a comunidade era composta por aproximadamente um milhão de contribuidores e de 6.033.568 usuários registrados (WIKI OSM, 2020), além de um grupo de aproximadamente 40 voluntários que dedicava seu tempo para criar e aperfeiçoar a infraestrutura do OSM, incluindo a manutenção dos servidores, a escrita do código do *software* que opera as transações com o servidor e a geração dos resultados cartográficos. Há também um grupo crescente de desenvolvedores de *software* que geram ferramentas para tornar disponíveis os dados do OSM para uso futuro em diferentes domínios de aplicação, plataformas e dispositivos (HAKLAY AND WEBER, 2008).

Os colaboradores do *OpenStreetMap* produzem os dados de diferentes modos: traçando as feições sobre imagens de satélite, importando dados de outros bancos de dados com licença pública, coletando dados através de dispositivos com *Global Positioning System* (GPS) enquanto caminham ou dirigem, preenchendo as etiquetas das feições ou, apenas corrigindo os nomes de uma rua ou localidade. Agrega-se ainda os dados fornecidos por agências oficiais de mapeamento e por outros produtores de dados geoespaciais. Os dados produzidos compõem um banco de dados geoespacial que pode gerar diferentes mapas além do mapa topográfico, como é o caso do *OpenCycleMap* um mapa específico para ciclistas. Além disso, os dados são revisados e corrigidos por dezenas de contribuidores e as atualizações podem ser visualizadas em minutos (WIKI OSM, 2020).

Os usuários também são convidados a adicionarem *tags* ou etiquetas aos dados que incluem na plataforma, com os nomes das feições e tipos de feições, entre outras informações. Estas *tags* funcionam como classes e atributos e fazem parte do modelo conceitual e semântico do OSM, o qual é descrito na próxima seção.

Alguns autores como (GRÖCHENIG, 2014; BENNETT, 2010) citam que o objetivo do *OpenStreetMap* é criar um banco de dados geoespacial de ruas (*street*) de todo o planeta. Mas, Steve Coast diz em seu livro “*The Book of OSM*” (COAST, 2015), que o *OpenStreetMap* (tradução livre) “é muito mais do que apenas ruas. Ele abrange qualquer coisa que você possa mapear, incluindo trilhas, restaurantes, trilhas de bicicleta, áreas de esqui, árvores, praias. Tudo que você imaginar, provavelmente alguém já o mapeou.” Na wiki do OSM (OSM, 2020), o *OpenStreetMap* é definido como “um projeto de mapeamento colaborativo para criar um mapa livre e editável do mundo”. Entretanto, as ruas ou vias têm papel relevante

no mapeamento da plataforma. De acordo com Board (1978), a navegação é uma das tarefas básicas de leitura de mapas utilizadas em aplicações geoespaciais. A navegação (Board, 1978) “é aplicada para facilitar o deslocamento de um lugar a outro” e três elementos são essenciais para a execução desta tarefa: a seleção da rota; a manutenção do curso e; o encontro do objetivo (BOARD, 1978). Sendo as ruas elementos essenciais à tarefa de navegação, a sua relevância se reflete tanto no nome da plataforma quanto na superioridade no número de vias mapeadas em relação a outros elementos no *OpenStreetMap*. No dia 28 de Abril de 2020, o número de caminhos (elemento geométrico utilizado para mapear as vias no OSM, o qual será tratado na próxima seção) mapeados era 659.492.871, superado apenas pelo número de pontos GPS que era de 7.807.397.088.

#### 2.5.2.1 Modelo Conceitual

O modelo conceitual do *OpenStreetMap* é baseado em *Extensible Markup Language* (XML).

A XML é uma linguagem de marcação e um padrão recomendado pelo *World Wide Web Consortium* (W3C) para a criação de documentos com dados organizados hierarquicamente. Esta linguagem é extensível porque permite definir os elementos de marcação. A linguagem de marcação é um agregado de códigos que podem ser aplicados a dados ou textos para serem lidos por computadores ou pessoas e tem como objetivo organizar, separar e integrar este conteúdo com outras linguagens (W3C, 2020). A XML também pode ser definida como um conjunto de regras para definir marcadores semânticos, que segmenta um documento em partes identificáveis. Trata-se de uma metalinguagem que define uma sintaxe para ser usada na geração de outras linguagens de marcação para domínios específicos com estrutura e semântica próprios (W3C, 2020).

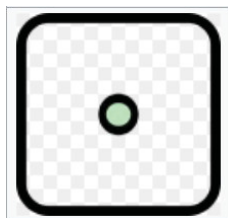
Como linguagem a *Extensible Markup Language*, torna possível o trabalho em conjunto com o padrão *Resource Description Framework* (RDF) fundamentado em triplas (sujeito, predicado e objeto) e outras linguagens como a *Web Ontology Language* (OWL) e o *Simple Knowledge Organization System* (SKOS) para atribuir semântica à representação de dados da *internet* (SOUSA et al., 2018).

A *OpenStreetMap* API é uma *interface Representational State Transfer* (*REST*) de serviço *Web* para ler e gravar no banco de dados usando o formato XML para transmitir dados por *Hypertext Transfer Protocol* (HTTP). São utilizados *Uniform Resource Locator* (*URLs*) para acessar objetos, e o padrão de código de resposta HTTP (WIKI OSM, 2020).

De acordo com (WIKI OSM, 2020b) o modelo conceitual do *OpenStreetMap*, é constituído de elementos simples, chamados "nós", "caminhos" e "relações", os quais têm equivalentes no banco de dados como primitivas gráficas. Cada elemento pode ter um conjunto de propriedades semânticas que são chamadas de "etiquetas" ou *tags*, por exemplo, *highway=primary*.

Um nó é o elemento básico do modelo do OSM, o qual consiste de latitude, longitude e opcionalmente altitude. Os nós devem obrigatoriamente ter uma etiqueta. Além disso, os nós são necessários para definir um caminho. O nó equivale à geometria ponto e a sua representação gráfica é mostrada na figura 17.

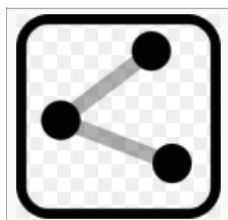
FIGURA 17 – NÓS NO OSM



FONTE: Wiki OSM (2020).

Um caminho é uma interconexão sem ordem, de no mínimo dois e no máximo dois mil nós. Um nó pode pertencer a vários caminhos. Os caminhos podem ser divididos em partes menores quando têm propriedades diferentes. Por exemplo, se uma seção de uma estrada contém um trecho de mão única, essa seção vai se tornar um caminho diferente a partir da seção de tráfego nos dois sentidos, mesmo que compartilhem o mesmo nome. O caminho equivale à geometria linha. A representação gráfica do caminho aberto é mostrada na figura 18.

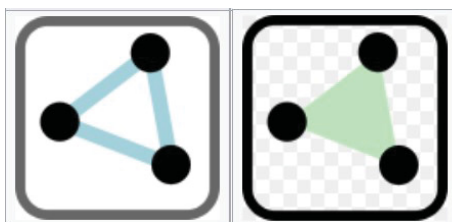
FIGURA 18 – CAMINHO ABERTO NO OSM



FONTE: Wiki OSM (2020).

As áreas são caminhos fechados. O primeiro e o último nós são coincidentes. Os caminhos fechados não necessariamente representam uma área. Este comportamento é delegado à interpretação das etiquetas. Para desenhar uma área dentro de outra área, é necessário recorrer a uma relação de multipolígono. Neste caso, um caminho fechado equivale à geometria área. As representações gráficas do caminho fechado e da área são mostradas na figura 19.

FIGURA 19 – CAMINHO FECHADO E ÁREA NO OSM



FONTE: Wiki OSM (2020).

Uma relação é uma estrutura de dados que documenta um relacionamento entre um ou mais elementos: nós, caminhos e/ou outras relações. Os elementos individuais são os “membros” da relação, que são usados para definir relacionamentos lógicos ou geográficos entre outros elementos e cada um deles tem uma função. Um membro de uma relação pode, opcionalmente, ter uma função que descreve a parte que um recurso particular desempenha dentro de uma relação. Por exemplo: uma restrição de circulação em uma via pode ter membros com as funções “daqui” e, “até aqui”. As relações podem ter diferentes significados, os quais são definidos através das etiquetas. Uma relação pode possuir um número arbitrário de etiquetas. Normalmente, espera-se que seja definida a chave *type=\**, que especifica o tipo da relação. A relação pode ocorrer em multipolígonos (um dentro do outro), rotas, restrições de acesso ou para definir exclusivamente limites administrativos

(WIKI OSM, 2020). A representação gráfica de membro, papel e relação são respectivamente mostrados na figura 20.

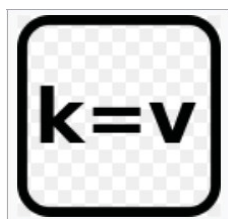
FIGURA 20 – MEMBRO, PAPEL E RELAÇÃO NO OSM



FONTE: Wiki OSM (2020).

As etiquetas desempenham a função semântica no modelo conceitual do OSM. As etiquetas consistem de dois campos de texto, uma *key* (chave) e um *value* (valor), que são sequências de caracteres *Unicode* com mais de 255 caracteres. Assim como os elementos (nós, caminhos e relações) implementam as feições, estes pares de "chave+valor", são elementos semânticos que descrevem o significado de cada feição adicionada ao mapa em forma similar ao de uma classe e atributo ou tipo de feição. Por exemplo, *amenity=restaurant* que corresponde a classe Edif\_Comerc\_Serv atributo Restaurante. Para preenchê-los, o OSM recomenda seguir seu padrão de rotulagem, documentado na página *Map Features* da sua *Wiki* (WIKI OSM, 2020). A representação gráfica das etiquetas no modelo conceitual do OSM é mostrada na figura 21.

FIGURA 21 – ETIQUETAS NO OSM



FONTE: Wiki OSM (2020).

O OSM tem toda a sua documentação disponibilizada através de *Wikis*. Uma *Wiki* é uma ferramenta de *software* ou um gerenciador de conteúdo que permite que qualquer pessoa crie e altere páginas de um *site* na *Web*. A principal característica desta ferramenta é a facilidade de edição e a possibilidade de criação



de textos de forma coletiva e livre. Tanto o modelo de dados quanto o modelo semântico, bem como tudo o que está relacionado ao *OpenStreetMap* é documentado pelos seus contribuidores detalhadamente e, em diversos idiomas na sua *Wiki*. As Wikis demonstram bem o modo de estruturação do modelo conceitual do OSM através da XML, na qual é possível navegar através de *links* por diversas páginas e conteúdos relacionados.

A figura 22 mostra a página da *Wiki* do OSM para a chave *highway* que serve para descrever as vias terrestres. Pode-se observar que a especificação é bem detalhada e apresenta: a chave, os possíveis valores para ela (os tipos de vias terrestres), a representação gráfica do elemento que deve ser utilizado para este tipo de feição, a descrição semântica da combinação chave-valor para a feição, o desenho ou a simbologia no mapa e ainda, uma foto da feição. Este nível de detalhamento supera o da ET-EDGV 3.0.

FIGURA 22 – DESCRIÇÕES SEMÂNTICAS VIAS TERRESTRES NO OSM

https://wiki.openstreetmap.org/wiki/Pt:Key:highway

### Vias terrestres

Etiquetas usadas para descrever estradas e caminhos terrestres. Veja também a página [vias terrestres](#) para outras orientações e [restrições](#) para detalhes sobre restrições de acesso por tipo de veículo/horário/dia/carga/propósito etc. As imagens apresentadas na tabela seguinte são apenas *alguns exemplos típicos*; não esgotam a variedade de exemplos que podem ser encontrados para uma mesma classe (também porque a classificação não é baseada unicamente em características físicas/visuais). No Brasil, certifique-se de verificar a [classificação viária brasileira](#).

Chave	Valor	Elemento	Descrição	Desenho	Foto
<b>Estradas</b>					
Estas são as principais etiquetas para a rede de estradas. Classificam-se desde a mais importante até a menos importante.					
highway	motorway		<b>Autoestrada</b> (ou oficialmente no Brasil: <b>rodovia</b> e por vezes como <b>via expressa</b> ) é uma via de acesso restrito a veículos motorizados e de alta velocidade, normalmente sob pagamento, com 2 ou mais faixas de trânsito e 1 faixa de acostamento/berma, com um separador central (raíles, blocos em cimento, simples faixa em terra), sem cruzamentos ou obstruções (semáforos, lombadas/lombas, etc.) Os cruzamentos com outras vias são desnivelados através de túneis ou pontes, por isso não são utilizados semáforos.		
highway	trunk		<b>Via expressa</b> <sup>(Brasil)</sup> ou <b>via rápida</b> <sup>(Portugal)</sup> : são estradas importantes que não são autoestradas. Geralmente têm separador central (opcional) e não são de acesso pago. Tal como as autoestradas são reservadas a veículos motorizados.) <ul style="list-style-type: none"> <li>Brasil: usar em <b>Itinerários Principais (IP)</b> e <b>Itinerários Complementares (IC)</b>.</li> <li>Portugal: usar em <b>Itinerários Principais (IP)</b> e <b>Itinerários Complementares (IC)</b>.</li> </ul>		
highway	primary		<b>Estrada primária</b> : liga grandes cidades, normalmente com duas pistas/faixas de rodagem, uma em cada sentido. Normalmente as faixas não têm barreira/separador central. <ul style="list-style-type: none"> <li>Brasil: usar em estradas federais e estaduais. Quando rural, costuma ser pavimentada, com apenas uma faixa por sentido e possui acostamento.</li> <li>Portugal: usar em <b>Estradas Nacionais (EN)</b>.</li> </ul>		

FONTE: Wiki OSM (2020).

Para cada chave e para cada combinação de chave+valor existe uma página na *Wiki* do OSM. Na página da chave, encontra-se a descrição da chave e de todas as etiquetas que ela contém, conforme é possível também observar na figura 22.

Por conseguinte, para cada uma das etiquetas também existe uma página na *Wiki* do OSM. Dentro dessas páginas específicas de cada etiqueta pode haver

sugestões de outras etiquetas que podem complementar a descrição da feição a que ela se refere, de modo similar aos atributos no modelo conceitual brasileiro. Existem duas possibilidades entre essas sugestões, uma delas é chamada de “combinações úteis” e a outra de “tags opcionais”. Por exemplo, para a etiqueta *man\_made=tower* que foi alinhada à classe torre de energia na ET-EDGV 3.0, as combinações úteis são mostradas na figura 23, são elas *tower:type=\** e *building=\**.

FIGURA 23 – COMBINAÇÕES ÚTEIS



FONTE: Wiki OSM (2020).

Dentro da mesma página específica da etiqueta existem ainda as “tags necessárias” ou “etiquetas necessárias” conforme mostra a figura 24. No caso, da combinação da etiqueta *tower:type=\** com a etiqueta de origem *man\_made=tower*, tem-se, por exemplo, *man\_made=tower* e *tower:type=communication* uma “etiqueta dupla” que melhor descreve a feição para fins de alinhamento semântico. Ou seja, tem-se uma torre e, o tipo da torre é de comunicação.




FIGURA 24 – ETIQUETAS NECESSÁRIAS

Necessário Tags

Key	Value	Element	Comment
man_made	tower	<input type="checkbox"/>	Um torre construído.
tower.type	(veja abaixo)	<input type="checkbox"/>	Além disso descrever o tipo de torre. Não é realmente obrigatório, mas é uma idéia muito boa para especificar um tipo.

Tipos Torre

Por favor, especifique o tipo de torre: motores de renderização pode usar as informações neste campo para mostrar ícones do mapa mais adequado ou irrelevante para ocultar tipos de torre.

Key	Value	Element	Comment	Example	Icon
tower.type	communication	<input type="checkbox"/>	Torres de comunicação que são utilizados para comunicações broadcast ou bidirecional via ondas de rádio, como rádio, televisão, telefones celulares e rádios bidirecionais. As antenas individuais / transponders podem ser etiquetados com <a href="#">Communications_Transponder (en)</a> .		
tower.type	observation	<input type="checkbox"/>	Torres de observação, tais como incêndio, torres, plataformas de observação e atrações turísticas.		
tower.type	<user-defined>	<input type="checkbox"/>	Se os valores acima não estão próximas o suficiente para o que você quer, ver <a href="#">"tower.type" no TagInfo (en)</a> , ou apenas fazer-se seus próprios valores. Por favor, documento recém-criada marca de valores aqui, de preferência, após <a href="#">discussão</a> .		

FONTE: Wiki OSM (2020).

Na mesma página desta etiqueta, existem também as “tags opcionais”, conforme ilustrado pela figura 25. Essas etiquetas opcionais funcionam de modo semelhante aos atributos no modelo conceitual brasileiro. Por que semelhante? Porque conforme é apresentado na seção da Metodologia mais adiante, o alinhamento entre os elementos conceituais e semânticos entre as duas fontes não é direto.

FIGURA 25 – ETIQUETAS OPCIONAIS

## Opcional Tags

Algumas marcas gerais que são usadas em muitos outros tipos de objeto são usadas para marcar torres de comunicação:

Key	Value	Element	Comment
name	*	<input type="checkbox"/>	Nome da torre
ref	*	<input type="checkbox"/>	Referência da torre
operator	*	<input type="checkbox"/>	Mantenedor / Operador da estrutura da torre.
height	*	<input type="checkbox"/>	Altura do topo da torre (incluindo as antenas) do nível do solo, em metros. No caso onde a torre está montado em cima de um prédio a altura deve incluir a altura do edifício.
ele	*	<input type="checkbox"/>	Elevação da base da torre acima do nível do mar em metros.
access	*	<input type="checkbox"/>	O acesso físico a torre, o mais provável privado, mas algumas torres têm visão plataformas.
source	*	<input type="checkbox"/>	Onde é que vêm os dados?
url	*	<input type="checkbox"/>	URL link para informações sobre este Tower.
tower:construction	lattice	<input type="checkbox"/>	A torre é construída em treliça de aço (a maioria tem cara fios)
tower:construction	freestanding	<input type="checkbox"/>	A torre é independente da construção "pesado", como concreto, aço ou madeira
tower:construction	dish	<input type="checkbox"/>	A "torre de comunicação "é um prato parabólico
tower:construction	dome	<input type="checkbox"/>	A "torre de comunicação ", é uma cúpula (ou "bola de golfe") de construção, com elementos de antena escondido da vista por uma <a href="#">radome</a> <sup>(en)</sup>
tower:construction	concealed	<input type="checkbox"/>	A "torre de comunicação " está escondido / disfarçado (por exemplo: feito para se parecer com uma árvore).

FONTE: Wiki OSM (2020).

A figura 26 mostra a área da plataforma *OSM*, no modo de edição, na qual os contribuidores inserem as informações semânticas para as feições adicionadas. É possível observar que a documentação também é detalhada, com descrições que orientam o contribuidor a preencher os campos, além de exemplos pictóricos.

FIGURA 26 – INSERÇÃO DAS INFORMAÇÕES SEMÂNTICAS NO OSM

The screenshot shows the 'OpenStreetMap' logo at the top left, with 'Editar' and 'Histórico' buttons. Below is a header 'Editar elemento' with a back arrow and a close 'X' button. The main content area is titled 'Via Expressa' with a road icon and an information icon. A dropdown menu 'Todos os campos' is expanded, showing three fields:

- Nome:** A text input containing 'Linha Verde'. Below it is a description: 'A etiqueta principal utilizada para atribuir um nome a um elemento.' and a link 'Ver no wiki do OpenStreetMap'. To the right is a thumbnail of a road sign.
- Número da Estrada:** A text input containing 'BR-476'. Below it is a description: 'Usado para indicar a referência ou código de algo como estradas, saídas de autoestradas, rotas, etc.' and a link 'Ver no wiki do OpenStreetMap'. To the right is a thumbnail of a road sign with 'N-330' and 'TERUEL'.
- Mão Única:** A checkbox labeled 'Sim' which is checked. To the right is a button 'Mudar Sentido <'. Below it is a description: 'Vias de sentido único/mão única são vias onde só se pode circular numa' and a link 'Ver no openstreetmap.org'. To the right is a thumbnail of a blue directional road sign.

FONTE: Contribuidores do *Open Street Map* (2020).

Na próxima seção são apresentadas as informações sobre a qualidade do mapeamento colaborativo relevantes para o contexto dessa pesquisa, bem como, estado da arte das pesquisas sobre esse assunto. Ao final da seção é apresentado um quadro com as reflexões sobre a qualidade na integração do mapeamento colaborativo ao mapeamento topográfico oficial.

#### 2.5.2.2 Qualidade

A qualidade dos dados do *OpenStreetMap* tem sido objeto de estudo de diversas pesquisas, entre elas as de (LUDWIG & ZIPF, 2019; TOUYA *et al.*, 2017; SIEBER & JOHNSON, 2013; ESTIMA e PAINHO, 2013; GOODCHILD & LI, 2012; BRONW, 2012; HAKLAY, 2010; FLANAGIN & METZGER, 2008; WILKINSON & HUBERMAN, 2007; BEARDEN, 2007).

Algumas pesquisas sugerem que a qualidade do VGI aumenta com o número de contribuidores, a mesma abordagem exemplificada na seção 2.5.1 sobre inteligência coletiva, e que (RAYMOND,1999) chamou de *Linu's Law* no contexto da engenharia de *software* (OLTEANU-RAIMOND, *et al.*, 2017; HAKLAY *et al.* 2010; GIRRES & TOUYA, 2010; FOODY, *et al.* 2015).

De acordo com Goodchild & Li (2012), se um contribuidor insere uma informação errada, outros contribuidores tendem a corrigi-lo e, o sucesso desse princípio aumenta à medida que o número de contribuidores também aumenta.

Possivelmente a heterogeneidade no mapeamento do OSM é a questão de maior preocupação para agências governamentais que estudam como utilizar estes dados para atualização cartográfica. Sobre esta questão Haklay (2010) observou a diferença existente entre o mapeamento do OSM em áreas urbanas e em áreas rurais. Os centros urbanos de grandes cidades da Inglaterra são bem mapeados, em contraste com as áreas rurais e periferias que recebem poucas contribuições.

Camboim, *et al.* (2015) observaram as mesmas discrepâncias na Região Metropolitana de Curitiba e analisaram a correlação entre os padrões de atualização dos dados do OSM com dados econômicos e sociais do censo do IBGE. Os autores chegaram à conclusão que as cidades nas quais os dados eram abundantes e atualizados frequentemente, eram as mesmas com mais recursos para investir em mapeamentos, indicando que existia correlação com fatores econômicos e sociais.

Alguns conjuntos de informações colaborativas podem ser bem detalhados porque foram importados de dados oficiais, por exemplo, e menos detalhados em algum lugar onde os dados foram obtidos através de imagens de satélite (OLTEANU-RAIMOND, *et al.*, 2017). De acordo com Rönneberg *et al.* (2019) os dados OSM são menos completos ou até ausentes em determinadas áreas, devido à distribuição geográfica dos colaboradores e porque se enquadram em diferentes tipos de colaboradores, que geram dados de qualidade variável.

A qualidade do VGI é frequentemente medida com relação a uma base cartográfica oficial, a qual pressupõe uma qualidade superior à do mapeamento colaborativo. Entretanto, muitas vezes essa base cartográfica inexistente, por isso procura-se inferir a qualidade através de propriedades intrínsecas como a densidade de feições, a quantidade de atualizações, a quantidade de contribuidores e o registro da origem dos dados, entre outros (CAMBOIM, *et al.*, 2015).

Com base em outras pesquisas (OLTEANU-RAIMOND, *et al.*, 2017) dizem que já foi demonstrado que detalhes chave de feições individuais podem ser inferidas de sua geometria, semântica ou metadados, permitindo que sua qualidade seja avaliada e possibilitando a integração destes dados aos mapeamentos oficiais.

Sob o aspecto semântico, uma questão importante, é que embora os conjuntos de dados oficiais tenham alto grau de confiabilidade, frequentemente carecem de informações semânticas que não sejam relativas aos propósitos administrativos e comerciais específicos aos quais servem. Por outro lado, as plataformas mapeamento colaborativo fornecem informações de modo fácil e livremente acessíveis sobre feições urbanas, com considerável grau de precisão geométrica e abrangência semântica (NOVACK *et al.*, 2018; COLEMAN, 2013).

Conforme explicado no início da seção 2.4.3.2 que trata da questão da qualidade no mapeamento topográfico oficial, não foi objetivo dessa pesquisa, avaliar a qualidade dos dados do mapeamento colaborativo em relação aos do mapeamento topográfico oficial brasileiro. Entretanto, o estudo da qualidade no contexto dos dois mapeamentos levou a algumas reflexões sobre a questão da qualidade na integração dos mapeamentos. O quadro 2 mostra uma síntese dessas reflexões (TOUYA, *et al.*, 2017; DSG, 2016b; IBGE, 2017).



QUADRO 2 – REFLEXÕES SOBRE A QUALIDADE ENTRE OS DOIS MAPEAMENTOS

**DIFERENÇAS NA AVALIAÇÃO DA QUALIDADE NO MAPEAMENTO TOPOGRÁFICO OFICIAL BRASILEIRO E NO MAPEAMENTO COLABORATIVO DO OSM**

	Mapeamento Topográfico Oficial	Informações Geográficas Voluntárias
<b>Qualidade Geral</b>	Homogênea	Heterogênea: granularidade das informações, qualidade varia de tema a tema, e também de feição a feição no mesmo tema.
<b>Mecanismos de Controle</b>	Normas, padrões e especificações	WikiOSM, metadados como histórico de edições, número de edições, reputação dos contribuidores.
<b>Compleitude</b>	São resultantes do processo de mapeamento adotado e podem ser avaliados e documentados de acordo com a ET-CQDG e o Manual de Avaliação da Qualidade do IBGE.	Problema dos espaços vazios, que podem ser realmente vazios ou apenas incompletos.
<b>Consistência Lógica</b>		Topologia das redes para identificar feições importantes como as rodovias p.ex, (largura das vias, etc.).
<b>Acurácia Posicional</b>		Feições incompatíveis no mesmo mapa, podem dar uma falsa ideia da realidade e do relacionamento entre as feições
<b>Acurácia Temporal</b>		Atualizações disponíveis em minutos.
<b>Acurácia Temática</b>		Problemas na classificação das feições, identificação depende da semântica, do contexto e da geometria.
<b>Escalas/Generalização</b>	Fixas e poucas.	Grande variedade de escalas, embora não haja generalização.
<b>Usabilidade</b>	Avalia se um produto atende às especificações dos usuários, por meio de eficiência, eficácia e satisfação em um contexto de uso específico.	Avalia se um produto atende às especificações dos usuários, por meio de eficiência, eficácia e satisfação em um contexto de uso específico.
<b>Level of Detail (LOD)</b>	É resultado do processo de mapeamento adotado e pode ser avaliado e documentado de acordo com a ET-CQDG e o Manual de Avaliação da Qualidade do IBGE.	Depende da resolução geométrica, precisão geométrica, granularidade e resolução semântica. É difícil inferir LOD como um valor numérico como para a escala, por isso são utilizadas categorias LOD para cidades, ruas, países, etc.

FONTE: A autora (2020).

Na próxima seção são apresentados os mecanismos de documentação dos metadados do *OpenStreetMap*, elementos que também permitem inferir a qualidade dos dados do mapeamento.

### 2.5.2.3 Metadados

Conforme citado anteriormente na seção 2.5.2.1 elementos como o histórico de edição, número de edições, comentários e número de edições do contribuidor são os metadados da plataforma *OpenStreetMap*. Estes metadados permitem inferir a qualidade dos dados da plataforma.

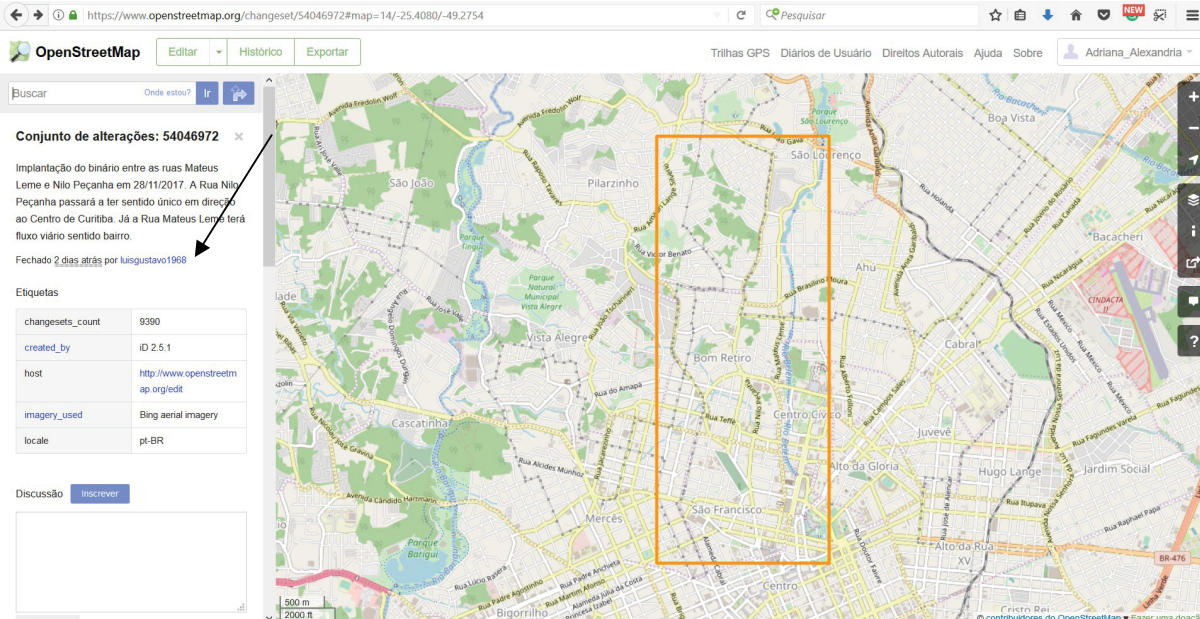
No “Histórico”, na barra superior do OSM é possível acessar o “Conjunto de Alterações” realizadas no mapa global e para uma determinada região, como mostra a figura 27. É possível observar que são descritos metadados como a natureza do conjunto de alterações, a data, as etiquetas ou *tags*, a fonte utilizada (imagem aérea



Bing), idioma e a identificação do contribuidor, além do polígono delimitador da área que sofreu as atualizações.

Na figura 28 é apresentado um detalhe deste conjunto de alterações mostrando os elementos alterados. Neste caso, a geometria de algumas linhas referentes às ruas e, em algumas relações referentes a rotas de ônibus.

FIGURA 27 – METADADOS DO CONJUNTO DE ALTERAÇÕES



OpenStreetMap

Buscar Onde estou? Ir

Conjunto de alterações: 54046972

Implantação do binário entre as ruas Mateus Leme e Nilo Peçanha em 28/11/2017. A Rua Nilo Peçanha passará a ter sentido único em direção ao Centro de Curitiba. Já a Rua Mateus Leme terá fluxo viário sentido bairro.

Fechado 2 dias atrás por luisgustavo1968

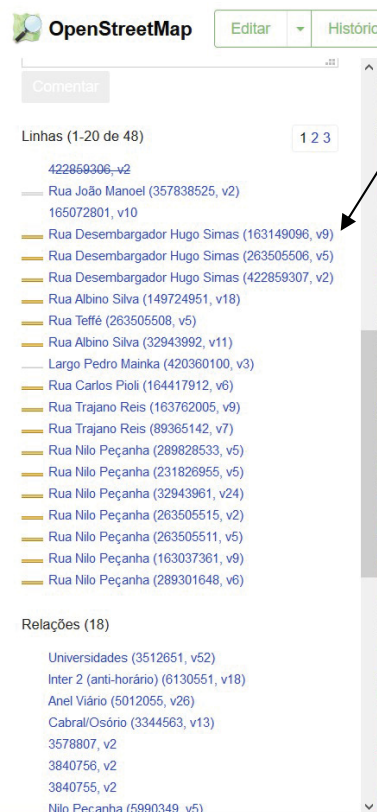
Etiquetas

changesets_count	9390
created_by	ID 2.5.1
host	http://www.openstreetmap.org/edit
imagery_used	Bing aerial imagery
locale	pt-BR

Discussão Inscrever

FONTE: Contribuidores do *Open Street Map* (2017).

FIGURA 28 – METADADOS DO CONJUNTO DE ALTERAÇÕES (cont.)



FONTE: Contribuidores do *Open Street Map* (2017).

Clicando no *link* que identifica cada uma das feições (pontos, linhas, relações) é possível acessar todo o histórico de edição da feição. Por exemplo, entre as linhas da figura 28, se escolhermos a Rua Desembargador Hugo Simas (como mostra a figura 29) e no final desta aba (figura 30) clicarmos em “Ver Histórico”, teremos acesso a todo o histórico de edição desta rua, como mostra a figura 31.

FIGURA 29 – METADADOS DA LINHA RUA DES. HUGO SIMAS

**OpenStreetMap** Editar Históri

Buscar Onde estou? Ir

**Linha: Rua Desembargador Hugo Simas (423879637)**

Adicionado localidades não demarcadas no mapa.

Editado 2 meses atrás por Rafael\_COE  
Versão #5 · Conjunto de alterações #54634207

**Etiquetas**

highway	secondary
name	Rua Desembargador Hugo Simas
oneway	no
source	yahoo_imagery

**Pontos**

370964457 (parte das linhas — Rua Desembargador Hugo Simas (29429441) e — Rua Santa Cecilia (171469876))

597427885 (parte da linha ..... Rua Afonso Pedri (46736915))

370964455

FONTE: Contribuidores do *Open Street Map* (2018).

FIGURA 30 – METADADOS DA LINHA RUA DES. HUGO SIMAS (cont.)

**OpenStreetMap** Editar Históri

Guilherme Seeger (32944270))

347153672

1632276020

347153644

370962252 (parte da linha — Rua Glácimo

Mylla (32944060))

370961447 (parte da linha — Rua Coronel João

Guilherme Guimarães (32943937))

347153622

1749662859

1749662900

347153606

370961380 (parte das linhas — Rua Ângelo Zeni (163215253) e — Rua Ângelo Zeni (171469877))

370962203 (parte das linhas — Rua Dom

Alberto Gonçalves (164416407) e — Rua Dom

Alberto Gonçalves (32944050))

347153596 (parte das linhas — Rua Carlos Pioli (32943948) e — Rua Carlos Pioli (163037375))

370961413 (parte das linhas — Rua Tapajós (420771365) e — Rua Henrique Itiberê da Cunha (163037373))

370961512

370961511

370961510 (parte das linhas — Rua Teffé (542568154), — Rua Teffé (543475642) e — Rua Desembargador Hugo Simas (543475639))

Baixar XML · Ver Histórico

FONTE: Contribuidores do *Open Street Map* (2018).

FIGURA 31 – HISTÓRICO DE EDIÇÃO DA RUA DES. HUGO SIMAS

**Histórico da linha: Rua Desembargador Hugo Simas (423879637)**

Adicionado localidades não demarcadas no mapa.

Editado 2 meses atrás por Rafael\_COE

Versão #5 - Conjunto de alterações #54634207

highway	secondary
name	Rua Desembargador Hugo Simas
oneway	no
source	yahoo_imagery

**Pontos**

370964457 (parte das linhas Rua Desembargador Hugo Simas (29429441) e Rua Santa Cecília (171469876))

597427885 (parte da linha Rua Afonso Pedri (46736915))

370962264 (parte da linha Rua Jornalista Octávio Secundino (32944063))

370962221 (parte da linha Rua Coronel Isaias Natel de Paula (32944053))

347153750 (parte da linha Rua Francisco Schaffer (31204350))

347153710

370964395 (parte da linha Rua Doutor Guilherme Seeger (32944270))

347153672

1632276020

347153644

370962252 (parte da linha Rua Glácomo Mylla (32944060))

370961447 (parte da linha Rua Coronel João Guilherme Guimarães (32943937))

347153622

1749662859

1749662900

347153606

370961380 (parte das linhas Rua Ângelo Zeni (163215253) e Rua Ângelo Zeni (171469877))

370962203 (parte das linhas Rua Dom Alberto Gonçalves (164416407) e Rua Dom Alberto Gonçalves (32944050))

347153596 (parte das linhas Rua Carlos Pioli (32943948) e Rua Carlos Pioli (163037375))

370961413 (parte das linhas Rua Tapajós (420771365) e Rua Henrique Itiberê da Cunha (163037373))

370962203 (parte das linhas Rua Dom Alberto Gonçalves (164416407) e Rua Dom Alberto Gonçalves (32944050))

347153596 (parte das linhas Rua Carlos Pioli (32943948) e Rua Carlos Pioli (163037375))

370961413 (parte das linhas Rua Tapajós (420771365) e Rua Henrique Itiberê da Cunha (163037373))

370961512

370961511

370961510 (parte das linhas Rua Teffé (542568154) e Rua Teffé (543475642) e Rua Desembargador Hugo Simas (543475639))

370961509 (parte das linhas Rua Desembargador Hugo Simas (543475639) e Shopping da Rua Teffé (171672304))

370961508 (parte da linha Rua Desembargador Hugo Simas (543475639))

370961507 (parte da linha Rua Desembargador Hugo Simas (543475639))

370961506 (parte da linha Rua Desembargador Hugo Simas (543475639))

370961505 (parte das linhas Rua Desembargador Hugo Simas (543475639) e Rua Desembargador Hugo Simas (418179636) e Rua Emilio de Menezes (418179635) e Rua Emilio de Menezes (163149093))

FONTE: Contribuidores do *Open Street Map* (2018).

Ainda observando a figura 31, se clicarmos em cada um dos pontos que fazem parte dessa linha, também será possível acessar os metadados com o histórico de cada uma das alterações nele ocorridas.

Clicando no *link* que identifica o nome do contribuidor (luisgustavo1988, na figura 27) é possível acessar uma página (figura 32) que mostra o número de edições realizadas por ele, as notas de mapa (notificações ou comunicações de erros que ele observou e, que outros contribuidores podem resolver), trilhas que ele inseriu, diário (que são informações que ele pode postar) e comentários que ele pode fazer nos diários de outros contribuidores.

FIGURA 32 – IDENTIFICAÇÃO DO CONTRIBUIDOR

**luisgustavo1968**

Edições 9.390 | Notas de Mapa | Trilhas 70 | Enviar Mensagem | Diário 0 | Comentários | Adicionar como Amigo

Mapeador desde: 12 de Junho de 2012

FONTE: Contribuidores do *Open Street Map* (2017).

A partir do *link* das edições é possível acessar todo o histórico de edições que este contribuidor realizou (figura 33).

FIGURA 33 – HISTÓRICO DE EDIÇÕES DO CONTRIBUIDOR



FONTE: Contribuidores do *Open Street Map* (2017).

À partir do *link* das trilhas é possível acessar as trilhas públicas do contribuidor realizadas por meio de *GPS* com os respectivos metadados (figura 34).



FIGURA 34 – TRILHAS PÚBLICAS GPS DO CONTRIBUIDOR



FONTE: Contribuidores do *Open Street Map* (2017).

As informações apresentadas demonstram que o histórico de atualização do *OpenStreetMap* contém os metadados descritos por geometria, feição, atributos, data, região e contribuidor, possibilitando rastrear de maneira bastante completa toda a sua origem e determinar a sua qualidade para os usos requeridos.

Um estudo feito por (BRAVO *et al.*, 2015) concluiu que os padrões de metadados adotados pela plataforma *OSM* são compatíveis com o padrão de metadados brasileiro, o Perfil MGB.

## 2.6 INTEGRAÇÃO DOS DADOS COLABORATIVOS AO MAPEAMENTO TOPOGRÁFICO OFICIAL

Nessa seção é apresentado um estudo de caso de integração de dados do mapeamento colaborativo da plataforma *OpenStreetMap* ao mapeamento topográfico oficial em grandes escalas do Instituto de Pesquisa e Planejamento Urbano de Curitiba (IPPUC).

Esse estudo foi realizado por ocasião do exame de qualificação e atende ao segundo objetivo específico desta pesquisa, que é realizar um teste prático de integração entre os bancos de dados do mapeamento colaborativo e do mapeamento topográfico oficial em grandes escalas de uma área de estudo, para testar as

potencialidades e desafios da compatibilização semântica para atualização e completude da base oficial. Agrega-se a este objetivo, o auxílio no levantamento de requisitos da ontologia de alinhamento entre os modelos conceituais, a obtenção de conhecimentos que propiciaram a geração do método de alinhamento semântico, bem como de conhecimentos para a geração do *plugin* desenvolvido no contexto dessa pesquisa, apresentado na seção da Metodologia.

O estudo de caso deu origem a um artigo científico publicado na Revista Brasileira de Gestão Urbana (URBE) no ano de 2019. O texto integral do artigo pode ser acessado e citado através do link <https://doi.org/10.1590/2175-3369.011.e20180142>.

Além de toda a descrição do estudo prévio de alinhamento semântico e integração dos mapeamentos, esse artigo contém a revisão de literatura referente ao estado da arte da integração do mapeamento colaborativo ao mapeamento topográfico oficial em diversos países e uma síntese com as potencialidades e os desafios relativos à integração.

### 3. METODOLOGIA

Para sistematizar o desenvolvimento das ontologias é necessário utilizar metodologias. Contudo, em Ciência da Computação existem poucas metodologias na literatura, as quais ainda não estão consolidadas (MORAIS & AMBRÓSIO, 2007). No contexto geoespacial, existem também poucas pesquisas com metodologias para a construção de ontologias. Os trabalhos não se detêm na metodologia, mas se referem a problemas específicos, como por exemplo, a construção de uma ontologia para aeroportos ou na definição de conceitos geoespaciais (PINHO & GOLTZ, 2003; SMITH & MARK, 2001; SMITH & MARK, 1998; SMITH, 1995). Entre as metodologias existentes e utilizadas em Ciência da Computação cita-se as descritas a seguir.

A *Enterprise*, que consiste de quatro fases: a identificação do propósito, a identificação do escopo, a formalização e a documentação formal. A *Methontology*, cujas etapas principais são: a especificação de requisitos; conceitualização do domínio do conhecimento; formalização do modelo conceitual em linguagem formal;

implementação de um modelo formal e manutenção da ontologia (incluindo a avaliação) (JONES *et al.*, 1998). A *On-To-Knowledge*, cujas fases são: o *kick-off* na qual além dos requisitos, são identificadas questões de competência, agregação de ontologias reutilizáveis e construída uma primeira versão da ontologia; o refinamento; a avaliação, onde são checados os requisitos iniciais e a ontologia é produzida efetivamente e; a manutenção, na qual são feitas adaptações e correção de eventuais erros (STAAB, 2001).

Uma última metodologia é a proposta por Falbo (1998), cuja abordagem é mais detalhada e que agrega as características das metodologias supracitadas já testadas na literatura. Essa metodologia cujo ciclo de vida é iterativo consiste de seis fases: (1) identificação de propósito e especificação de requisitos, (2) captura da ontologia, (3) formalização da ontologia, (4) integração com ontologias existentes, (5) avaliação da qualidade e (6) documentação da ontologia.

Esta metodologia (Falbo, 1998) foi utilizada para construir a ontologia que alinha semanticamente as ontologias do mapeamento colaborativo da plataforma OSM e do mapeamento topográfico oficial do Brasil, proposta por esta pesquisa.

De acordo com as informações apresentadas no capítulo 2 os modelos conceituais do mapeamento topográfico oficial e do mapeamento colaborativo foram considerados ontologias. O alinhamento entre essas ontologias produziu uma terceira ontologia e, a documentação de todo o processo de geração dessa ontologia produziu uma base de conhecimento.

A figura 35 apresenta o fluxograma com as etapas da metodologia utilizada nesta pesquisa.



FIGURA 35 – PROCESSO METODOLÓGICO



FONTE: A autora (2020).

Nas próximas seções são apresentadas as etapas metodológicas e o desenvolvimento de cada uma delas. Os resultados correspondentes são apresentados no capítulo seguinte Resultados e Discussões.

### 3.1 IDENTIFICAÇÃO DE PROPÓSITO E ESPECIFICAÇÃO DE REQUISITOS

A identificação de propósito e especificação de requisitos é a primeira etapa na construção da ontologia. A finalidade desta etapa é identificar as questões que a ontologia deve responder, ou seja, a competência da ontologia (FALBO, 1998).

O levantamento de requisitos no contexto da ciência da geoinformação não difere do levantamento da engenharia de *software* ou ciência da computação. Devem ser considerados não como uma diferença, mas como aspectos relevantes para o contexto da geoinformação, aqueles já discutidos anteriormente nas seções de interoperabilidade entre ontologias e componentes das ontologias. A saber, a localização espacial, os atributos espaciais e o tempo (SLUTER *et al.*, 2016).

De acordo com Sluter *et al.* (2016) “uma solução de geoinformação pode ser projetada e construída como um sistema destinado a representar e disseminar geoinformação”. Um sistema de geoinformação é considerado desde um mapa estático até um sistema complexo para visualização e análise de dados geoespaciais (SLUTER *et al.*, 2016).

Nessa etapa foram então identificados seus usos e propósitos, aspectos relevantes para a ontologia, seus potenciais usuários e o contexto que promoveu a sua criação. Como o autor Falbo (1998) sugere, esta metodologia tem ciclo de vida iterativo, então essa etapa foi recorrente durante todo o processo de alinhamento entre as ontologias.

A identificação destes aspectos foi compilada em uma tabela descrita na seção 4.1 do capítulo de Resultados e Discussões.

### 3.2 CAPTURA DA ONTOLOGIA

A segunda e mais importante etapa, é a de captura da ontologia. Com base nas questões de competência levantadas na etapa de identificação de propósito e especificação de requisitos, o objetivo desta etapa é identificar os conceitos, classes, relacionamentos, propriedades, axiomas e instâncias da ontologia (FALBO, 1998).

Para facilitar a comunicação com outros especialistas de domínio é recomendada a geração de um dicionário de termos e que a especificação seja também representada através de linguagem gráfica. Recomenda-se ainda, o uso de taxonomias para a representação dos conceitos. Conceitos que possam não ser entendidos a partir dos conceitos primitivos da ontologia devem ser especificados através de linguagem natural, evitando ambigüidades e inconsistências no entendimento dos termos (FALBO, 1998).

De acordo com Morais & Ambrósio (2007), a base de uma ontologia são os conceitos, suas propriedades e seus relacionamentos. A definição da semântica dos termos de uma ontologia é construída através de axiomas, os quais modelam sentenças sempre verdadeiras. Os axiomas individualizam as definições dos termos e as restrições sobre os seus significados e normalmente são especificados em linguagem natural. Falbo (1998), diz que os axiomas também podem ser especificados através de lógica de primeira ordem (LPO), uma linguagem utilizada para a formalização de ontologias computacionalmente.

Existem dois tipos de axiomas: os axiomas de derivação, que possibilitam especificar informações a partir de conhecimento preexistente e; os axiomas de consolidação, que definem a coesão entre as informações, como por exemplo, a definição de um elemento como instância de um conceito dentro do domínio. Os axiomas de derivação, por sua vez, são divididos em epistemológicos e ontológicos. Aqueles que tratam da estruturação dos conceitos e das suas relações são chamados de axiomas epistemológicos e, aqueles que tratam da interpretação dos termos e das suas restrições, bem como das leis de integridade que os regem são chamados de axiomas ontológicos (FALBO, 1998). Morais & Ambrósio (2007) citam que os axiomas devem ser necessários e suficientes para expressar as questões de competência da ontologia, bem como, para descrever as soluções para tais questões, sendo consistente com elas.

A definição dos conceitos, classes, relacionamentos, propriedades, axiomas e instâncias já havia sido realizada pelas instituições produtoras através dos modelos conceituais, os quais foram detalhados nas seções sobre a ET-EDGV 3.0 e sobre o modelo conceitual do OSM. Com base nos modelos conceituais, nessa etapa foi realizada então, a correspondência semântica de cada um dos elementos supracitados entre os esquemas conceituais, bem como, a documentação deste processo através de linguagem natural e gráfica.

Esta etapa metodológica foi subdividida em quatro fases: (1) obtenção das informações semânticas para o alinhamento; (2) estrutura dos modelos conceituais; (3) seleção das chaves do OSM e categorias da ET-EDGV; (4) método proposto para o alinhamento semântico. A fase (1) não tem um resultado na seção de Resultados e Discussões.

Nas próximas seções são descritas cada uma destas fases do processo de captura da ontologia, as quais atendem aos objetivos específicos 3, 4 e 5 propostos para esta pesquisa.

### 3.2.1 Obtenção das informações semânticas para o alinhamento

As informações da ontologia do mapeamento do *OpenStreetMap* foram obtidas na *wiki* do OSM (WIKI OSM, 2020). As informações da ontologia do mapeamento brasileiro foram obtidas na Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV 3.0) (DSG, 2017). Nos casos em que houve dúvidas na classificação, também foram consultadas as informações da Especificação Técnica para a Aquisição de Dados Geoespaciais Vetoriais (ET-ADGV 3.0) (DSG, 2018) e Manual Técnico T 34-700 Convenções Cartográficas (1ª Parte) Normas para o Emprego dos Símbolos (DSG, 1998).

Para as informações referentes às vias terrestres e à infraestrutura rodoviária também foram consultadas as normas, os manuais técnicos e a legislação pertinente listados a seguir: Manual de Pavimentação (DNER, 1996); Glossário de Termos Técnicos Rodoviários (IPR-700) (DNIT, 2017; Manual de Pavimentação (IPR-719) (DNIT, 2006); Manual de Implantação Básica de Rodovia (IPR-742) (DNIT, 2010); Terminologias Rodoviárias Usualmente Utilizadas (DNIT, 2007); Nomenclatura das Rodovias Federais (DNIT, 2020); Lei nº 9.503 que institui o Código de Trânsito Brasileiro (BRASIL, 1997) e; Lei nº 12.379 que dispõe sobre o Sistema Nacional de Viação (SNV) (BRASIL, 2011).

Para as informações referentes às edificações de comércio, serviços e industriais também foram consultados o Manual de Classificação Nacional de Atividades Econômicas (CNAE) (IBGE, 2015) e a Lei Complementar nº 116 (BRASIL, 2003) que dispõe sobre o imposto sobre serviços de qualquer natureza. Para todas as informações também foram consultados dicionários, *wikis*, pesquisa de imagens e notícias.

Esta etapa não tem um resultado direto, desse modo não há uma seção correspondente no capítulo de Resultados e Discussões.

### 3.2.2 Estrutura dos modelos conceituais

Na segunda fase da etapa de captura da ontologia foram analisadas as estruturas dos modelos conceituais dos dois mapeamentos. Uma consequência decorrente do fato de os modelos conceituais terem diferentes estruturas, é que a hierarquia das categorias, classes, atributos e valores de domínio da ET-EDGV 3.0 não têm correspondência direta com as etiquetas (chave+valor) do modelo conceitual do *OpenStreetMap*. Por isso, na etapa de captura da ontologia é recomendado o uso de taxonomias para a representação dos conceitos e também que a especificação seja representada através de linguagem gráfica.

Desse modo, para alinhar a estrutura conceitual e semântica dos dois modelos conceituais foram produzidas representações gráficas cujos resultados são apresentados na seção 4.2.1.

### 3.2.3 Seleção das chaves do OSM e categorias da ET-EDGV 3.0

Para realizar o alinhamento entre as ontologias, inicialmente foram selecionadas as chaves *amenity* e *highway* do mapeamento colaborativo do OSM. A chave *amenity* comporta as categorias de Edificações, Cultura e Lazer e Estrutura Econômica. A chave *highway* comporta a categoria Sistema de Transporte/Rodoviário. A escolha dessas chaves foi feita com base no estudo prévio realizado por Bortolini *et al.* (2018), que teve como objetivo levantar quais categorias teriam maior potencial de integração sob o aspecto legal. Segundo os autores (Bortolini *et al.*, 2018) as feições de Sistema de Transporte são muito difundidas no *OpenStreetMap* e podem contribuir com grande quantidade de informações geográficas com um processo de validação. Já as Edificações possuem uma grande e variada quantidade de feições sem a necessidade de um processo complexo de validação.

Todavia, como o modelo conceitual do OSM tem uma estrutura semântica menos rígida que a da ET-EDGV 3.0 e porque os modelos conceituais têm estruturas diferentes, ao longo do processo foram agregadas outras chaves do mapeamento colaborativo e também outras categorias da ET-EDGV 3.0. Para melhor compreensão cita-se como exemplo a chave *amenity* que, embora tenha a

maior parte das feições correspondentes às categorias Edificações, Cultura e Lazer e Estrutura Econômica, contém também a etiqueta *amenity=bus\_station*, que na ET-EDGV 3.0 corresponde a uma classe da categoria Sistema de Transporte. Outro exemplo, é a classe *Via\_Deslocamento* da ET-EDGV 3.0 que tem os atributos *revestimento* e *tipoPavimentacao* os quais utilizam a chave *surface=\**, levando a agregação também desta chave na compatibilização. O asterisco significa que existem diversos valores associados a essa chave, os quais assumem diferentes significados conforme a combinação de chave+valor. Um terceiro caso é o das “combinações úteis” do modelo do OSM, que são chaves que funcionam de modo semelhante aos atributos na ET-EDGV 3.0, as quais são sugeridas como complemento para cada chave.

Como resultado para esta etapa foi gerada uma tabela que apresenta a o número e a porcentagem de classes de cada categoria da ET-EDGV 3.0 alinhadas com as respectivas chaves correspondentes no *OpenStreetMap*, bem como, o número total de chaves alinhadas com a categoria em questão. Essa tabela e a análise dos resultados nela descritos são apresentados na seção 4.2.2.

### 3.2.4 Método proposto para o alinhamento semântico

As medidas de similaridade entre diferentes fontes de dados podem ser espaciais, entre termos ou séries de termos (*string*) e, semântica. A medida de similaridade espacial se refere à distância espacial (euclidiana) entre dois termos de diferentes conjuntos de dados. A medida de similaridade entre termos diz respeito à análise de combinações de termos (sinônimos, acrônimos, abreviações, pronomes, etc.). As medidas de similaridade semântica comparam a similaridade entre os conceitos (NOVACK *et al.*, 2018; AL-BAKRI e FAIRBAIRN, 2012).

Na literatura, são citados três métodos para comparar a similaridade semântica entre os conceitos. Esses métodos podem se basear em uma taxonomia ou modelo, em um conjunto de textos, ou em ambos. Quando o método é fundamentado em uma taxonomia ou modelo, diz-se que as medidas de similaridade são baseadas em conhecimento. Quando método se apoia em um conjunto de textos, diz-se que as medidas de similaridade são baseadas em um *corpus*. Quando

o método é respaldado por ambas as abordagens anteriores, diz-se que as medidas de similaridade são híbridas.

Existem alguns algoritmos e plataformas para medir as similaridades semânticas automaticamente como o *Shortest Path*, *Wup*, *Resnik*, *Lin*, *Jcn* e alguns corpora como o *Wordnet*, *Brown* e *SemCor* (todos em língua inglesa), entretanto a análise dos resultados sempre vai depender da avaliação de um especialista de domínio (NOVACK *et al.*, 2018; NOVACK *et al.*, 2016; BALLATORE *et al.*, 2014; AL-BAKRI e FAIRBAIRN, 2012).

No alinhamento semântico proposto por essa pesquisa foi utilizado o método híbrido, ou seja, com base na estrutura dos modelos conceituais e nas suas descrições textuais e de outras fontes, de modo manual.

Para propor um método para realizar o alinhamento semântico foi necessário ter compreendido e esquematizado a estrutura conceitual das ontologias do OSM e da ET-EDGV 3.0, pois como foi explicado anteriormente, devido às suas diferentes estruturas, os elementos e termos referentes aos conceitos não são diretamente alinhados. Desse modo, utilizando a estrutura das duas ontologias e as suas descrições textuais, foi proposto então o método para o alinhamento semântico. A figura 36 mostra as fases do método de alinhamento.

FIGURA 36 – MÉTODO PROPOSTO PARA O ALINHAMENTO SEMÂNTICO

**OSM (19/04/2019 a 06/11/2019) ➡ ET-EDGV 3.0 (homologada 12/2017)**

1. TERMOS DOS CONCEITOS	2. TERMOS DAS DESCRIÇÕES	3. ANÁLISE TEXTUAL	4. CONSULTA A OUTRAS FONTES	5. BUSCA CONCEITOS NA ET-EDGV
Verificação de termos iguais, semelhantes e sinônimos nos conceitos das duas ontologias.	Verificação de termos iguais, semelhantes e sinônimos nas descrições semânticas referentes aos conceitos.	Análise textual dos termos e descrições semânticas referentes aos conceitos.	Nos casos em que houve dúvidas foram consultadas outras fontes textuais e imagens.	Busca dos conceitos das etiquetas do OSM nas categorias, classes, atributos e valores de domínio na ET-EDGV 3.0.

FONTE: A autora (2020).

Primeiramente, foram considerados os termos dos conceitos e feições que fossem iguais, sinônimos ou semelhantes. Por exemplo, estrada, autoestrada, rodovia. Posteriormente, ou no caso de somente o termo referente à feição ou conceito não atender os requisitos da primeira seleção, foram então consideradas as



palavras, sinônimos ou semelhantes nas descrições semânticas referentes aos termos. Caso os requisitos do primeiro e segundo critérios de seleção não fossem atendidos ou parcialmente atendidos, então era feita a análise textual das descrições dos conceitos em ambas as fontes. Se houvesse dúvidas eram consultadas outras fontes, inclusive imagens, até que fosse encontrado o significado similar ou equivalente para efetuar a correspondência semântica. Por fim, eram feitas as buscas pela categoria, classe, atributo ou valor de domínio correspondente na ET-EDGV. Nos casos em que elementos da ET-EDGV não estavam na chave analisada no momento, mas que pudessem ter correspondente no *OSM* era também feita uma busca em outras chaves do *OpenStreetMap* para encontrar uma correspondência.

O alinhamento semântico foi feito do *OSM* para a ET-EDGV, tratando primeiro a chave (*key*) e em seguida cada um dos pares de chave+valor, ou seja, as chamadas etiquetas (*key+value=tag*) dela constantes. No processo houve empenho em compatibilizar o máximo possível as feições, visando à completude e a atualização do mapeamento nacional.

O processo de alinhamento foi realizado no período de 19 de Abril de 2019 a 06 de Novembro de 2019. Essa informação é importante porque as etiquetas do *OSM* estão em constante discussão e atualização pelos contribuidores da plataforma. De modo semelhante, a versão da ET-EDGV (3.0) também é relevante, pois a especificação passa por atualizações, as quais são validadas por uma comunidade reduzida, composta por especialistas de domínio. A ET-EDGV 3.0, ou seja, a 3ª edição da especificação foi homologada pela CONCAR no mês de dezembro de 2017. Embora menos freqüentes, estas atualizações podem gerar alterações substanciais, como no caso da ET-EDGV 3.0 em relação à versão anterior a ET-EDGV Defesa Força Terrestre (ET-EDGV F Ter) (DSG, 2016).

Por isso as ontologias devem ser documentadas e ter um processo de manutenção (NANVAR, *et al.*, 2019; ARENHART & KRAUSE, 2009; STAAB, 2001; JONES *et al.*, 1998). Um exemplo, desse tipo de alteração ocorreu durante o desenvolvimento dessa pesquisa. No estudo de caso realizado para a qualificação, foi utilizada a ET-EDGV F Ter (DSG, 2016). Na ocasião, houve dúvida na classificação de trecho rodoviário quando esse passava dentro do perímetro urbano, se deveria ser classificado como Trecho\_Arruamento ou Trecho\_Rodoviario. Na época foi decidido classificar esse trecho rodoviário dentro do perímetro urbano como trecho de arruamento. Na versão 3.0 da ET-EDGV esse impasse foi resolvido



com a inserção de um atributo *booleano* chamado *trechoEmPerimetroUrbano*. E, além disso, tanto o *Trecho\_Arruamento* quanto *Trecho\_Rodoviario* estão hierarquicamente subordinadas a uma classe genérica chamada *Via\_Deslocamento* na ET-EDGV 3.0, o que não ocorria na versão anterior.

Algumas classes foram diretamente alinhadas através dos termos e das descrições semânticas constantes das ontologias, como é o exemplo da classe *Arvore\_Isolada* da ET-EDGV 3.0 (DSG, 2017) e da chave *natural=tree* (OSM, 2020):

“**Árvore isolada**, no contexto desta especificação, é aquela que ocorre em espaços públicos (incluindo os trechos de arruamento), cuja responsabilidade pela administração cabe ao poder público. Em situações especiais representa também as árvores isoladas localizadas em áreas rurais sem presença de outras vegetações de grande porte, se apresentando assim, como ponto de referência para região”. (DSG, 2017)

“**Árvore**: uma única árvore. Esta etiqueta foi originalmente criada apenas para árvores de valor histórico ou cultural, mas atualmente é aplicada em qualquer tipo de árvore, pelo menos em zonas urbanas, servido para inventário e análise, sinalização de obstáculos para cegos, etc.” (OSM, 2020)

Contudo, em alguns casos, como o dos paraciclos ou estacionamentos de bicicletas, que tem a etiqueta *amenity=bicycle\_parking* no OSM, o alinhamento demandou uma análise mais aprofundada e não apenas dos termos e descrições dos conceitos conforme o método proposto nesta pesquisa. Esta feição ainda não existe na ET-EDGV 3.0, mas conforme explicado anteriormente, no processo de alinhamento das ontologias visou-se maximizar o aproveitamento das informações colaborativas disponíveis, para atualização e completude da base oficial. Então foi necessário buscar outras fontes, como por exemplo, o *site* da prefeitura de Curitiba, que foi utilizado no estudo de caso, o qual tem definições dos elementos da estrutura ciclovária da cidade (CURITIBA, 2017), a *wiki* do OSM que tem as descrições semânticas das *tags* (WIKI OSM, 2017) e imagens na *internet* para estabelecer o significado da feição. Em seguida foi necessário analisar o diagrama de classes da categoria Transportes para selecionar uma classe com a qual fosse possível alinhar o conceito de paraciclo, dada a sua importância para o contexto urbano atual.

Durante o alinhamento também surgiram algumas dúvidas conceituais em relação a determinados termos, como no caso de rodovia e estrada, que do ponto de vista do leigo, são utilizados como sinônimos. Tanto na estrutura conceitual quanto

nas descrições da ET-EDGV 3.0 e nas descrições do *OSM* as diferenças entre esses termos pode levar a interpretações ambíguas. De acordo com o autor da metodologia utilizada para essa ontologia de alinhamento, Falbo (1998), conceitos que possam não ser entendidos a partir dos conceitos primitivos da ontologia devem ser especificados através de linguagem natural, evitando ambigüidades e inconsistências no entendimento dos termos. Desse modo, foi considerada importante uma análise e documentação da situação.

Na estrutura conceitual da ET-EDGV 3.0 a classe *Via\_deslocamento* é definida como “a via de trânsito terrestre, destinada a veículos automotores, exceto o caminho carroçável e aqueles pertencentes ao Sistema Ferroviário” (DSG, 2017). Essa classe foi alinhada com a chave *highway* do *OSM* que é definida da seguinte forma “a etiqueta *highway* (rodovia) serve para descrever o tipo de via, como uma autoestrada ou um caminho pedonal” (OSM, 2020).

A classe *Via\_deslocamento* tem como atributo o *tipoVia* que “indica o tipo de via de deslocamento” (DSG, 2017), o qual por sua vez tem como valores de domínio Autoestrada e Rodovia, entre outros valores. A descrição semântica de Autoestrada é a seguinte (DSG, 2017):

“Via de tráfego rápido, com todos os acessos controlados, sem cruzamento de nível e destinada exclusivamente a veículos motorizados, com revestimento sólido (asfalto, concreto ou calçamento), com um mínimo de quatro faixas, apresentando separação física entre as pistas de tráfego, representável em escala ou não.”

No *OSM* a etiqueta correspondente é *highway=motorway* cuja descrição é (OSM, 2020):

“Autoestrada (ou oficialmente no Brasil: rodovia e por vezes como via expressa) é uma via de acesso restrito a veículos motorizados e de alta velocidade, normalmente sob pagamento, com 2 ou mais faixas de trânsito e 1 faixa de acostamento/berma, com um separador central (raíles, blocos em cimento, simples faixa em terra), sem cruzamentos ou obstruções (semáforos, lombadas/lombas, etc.) Os cruzamentos com outras vias são desnivelados através de túneis ou pontes, por isso não são utilizados semáforos.”

A descrição semântica de Rodovia na ET-EDGV é (DSG, 2017) “via destinada principalmente ao tráfego de veículos providos de pneus, que atravessa uma extensão territorial, ligando duas ou mais localidades”.

No OSM uma das etiquetas correspondentes ao atributo Rodovia na ET-EDGV 3.0 é *highway=trunk* cuja descrição é (OSM, 2020): “Via expressa (Brasil) são estradas importantes que não são autoestradas. Geralmente têm separador central (opcional) e não são de acesso pago. Tal como as autoestradas são reservadas a veículos motorizados.

Embora as descrições da ET-EDGV 3.0 sejam diferentes para autoestrada e rodovia não são muito esclarecedoras com relação à diferença entre elas. Quando se analisa as descrições das etiquetas correspondentes no OSM a situação também não é muito diferente. Por isso, foram consultados os documentos técnicos do Departamento Nacional de Infraestrutura Terrestre (DNIT), Departamento Nacional de Estradas e Rodagem (DNER) e Código de Trânsito Brasileiro (CTB) (BRASIL, 1997).

O Código de Trânsito Brasileiro (CTB) (BRASIL, 1997) diferencia estrada como “via rural não pavimentada” e rodovia como “via rural pavimentada”. O Glossário de Termos Técnicos Rodoviários do DNIT (IPR-700) traz as seguintes definições (DNIT, 2017):

**“ESTRADA** 1) Via de trânsito em zonas não urbanas. 2) Via de trânsito, em geral em zona não urbana, destinada a veículos rodoviários, animais e pessoas e que não tem as características de estrada de rodagem (**rodovia**), nem de auto-estradas. **V. Estrada de Rodagem e V. Auto-estrada.** 3) Termo genérico para designar via terrestre, inclusive estrada de rodagem (**rodovia**) e auto-estrada.

**Espanhol** – *Vía, Camino, Carretera, Senda (Nic.), Vereda (R. D.)*

**Francês** – *Chemin, Route, Voie Routière*

**Inglês** – *Road*

**ESTRADA DE RODAGEM** Estrada que, tendo a sua plataforma devidamente preparada e pavimentada, se destina à circulação de veículos automotores. Sinônimo: **Rodovia**.

**E** – *Vía, Carretera, Ruta*

**F** – *Autoroute, Route*

**I** – *Highway*

**AUTO-ESTRADA** Via de tráfego rápido, com todos os acessos controlados, sem cruzamento de nível e destinada exclusivamente a veículos motorizados. (Sinônimo: Estrada Bloqueada, **Rodovia Bloqueada**).

**E** - Autopista, Supercarretera, Auto-Ruta (Nic.)

**F** - Autoroute

**I** - Freeway, Motorway

**RODOVIA** V. Estrada de Rodagem.

**E** - *Vía, Carretera, Camino, Ruta (Pan.)*

**F** - Route

**I** - Highway, Road”

Embora a IPR-700 esteja em concordância com o CTB na primeira definição de estrada como rural, se contradiz ao utilizar rodovia como sinônimo na terceira definição e nas definições dos demais termos sugeridos, incluindo a definição de rodovia. Ou seja, estrada e rodovia são mais dois termos técnicos usados como sinônimos por especialistas de domínio, demonstrando que ainda são necessários muitos estudos, pesquisas semânticas e ontologias para esclarecer determinados conceitos. Dentre essas definições teve grande valia no alinhamento semântico desses dois termos, o fato de a IPR trazer nas definições também as palavras em outros idiomas. Entretanto, este tipo de descrição é ambígua, situação que gera problemas no alinhamento semântico, conforme foi constatado nos testes realizados (os testes são apresentados no capítulo de Resultados e Discussões na seção de Avaliação da Qualidade da Ontologia).

Situações semelhantes também ocorreram em relação aos conceitos dos atributos Revestimento e tipoPavimentação e, nos valores de domínio Ligação entre Pistas e Trecho de Entroncamento do atributo TipoVia também da classe Via\_deslocamento, entre outros.

Para esta fase da etapa de captura da ontologia foram gerados três resultados. O primeiro deles foi um conjunto de tabelas que contêm o alinhamento semântico propriamente dito, subdividido pela divisão de escalas do mapeamento (MapTopGE e MapTopPE) e categoria da ET-EDGV 3.0 e pela chave principal do modelo conceitual do OSM. O segundo resultado é uma tabela que apresenta os possíveis casos de correspondência semântica e geométrica entre as ontologias dos dois mapeamentos. O terceiro resultado gerado pelo alinhamento semântico entre as ontologias do OSM e da ET-EDGV 3.0 foi um diagrama na forma de grafos representando o alinhamento da chave *highway* do OSM com a categoria Sistema de Transporte e Sistema de Transporte/Rodoviário da ET-EDGV 3.0. Ambos os resultados são apresentados na seção 4.2.3.

Nas três próximas seções, são apresentadas as etapas finais da metodologia utilizada para a construção da ontologia de alinhamento entre o mapeamento topográfico oficial brasileiro e o mapeamento colaborativo do OSM.

### 3.3 FORMALIZAÇÃO DA ONTOLOGIA

A terceira etapa é a de formalização da ontologia, na qual a ontologia é representada através de uma linguagem computacional, considerada formal. Na prática, uma ontologia pode ser representada através de qualquer linguagem não formal, ou seja, através de linguagem natural, embora a representação formal seja considerada a mais apropriada. A escolha do *software* e/ou da linguagem depende do tipo e objetivo da ontologia a ser criada (MORAIS & AMBRÓSIO, 2007).

Do mesmo modo como ocorre com as metodologias, ainda não existe distinção entre as tecnologias utilizadas para a implementação das ontologias em Ciência da Computação e Ciência da Geoinformação.

Existem várias ferramentas para a construção de ontologias, contudo, de acordo com alguns autores, a manipulação dessas ferramentas não é uma tarefa trivial para usuários não-especialistas da área de informática (AL-BAKRI & FAIRBAIRN, 2012; PINHO & GOLTZ, 2003).

O principal *software* utilizado para a criação de ontologias é o *Protégé* (PROTÉGÉ, 2020), desenvolvido pelo grupo de informática médica da Universidade de *Stanford*. Trata-se de “uma plataforma gratuita e de código aberto que contém um conjunto de ferramentas para construir modelos de domínio e aplicativos baseados em conhecimento com ontologias”. Atualmente, está na versão 5.5.0 e suporta as especificações de linguagem ontológica da *Web* OWL 2 e RDF do *World Wide Web Consortium* (W3C) como, RDF/XML, Turtle, OWL/XML, OBO, OIL, Prolog e outros formatos disponíveis para *upload* e *download* de ontologias (PROTÉGÉ, 2020).

Outra plataforma utilizada para a criação de ontologias é a Ontolingua. Desenvolvida e mantida pelo *Knowledge Systems Laboratory* da Universidade de Stanford, a Ontolingua fornece um ambiente colaborativo e distribuído na *internet* (disponível em <http://www.ksl.stanford.edu/software/ontolingua/>) que permite navegar, criar, editar, modificar e usar ontologias (KSL, 2020). De acordo com Gruber (1993) a sintaxe e a semântica das definições da Ontolingua são baseadas em uma notação estendida da lógica de predicados de primeira ordem (LPO) chamado *Knowledge Interchange Format* (KIF). As definições da Ontolingua são escritas em linguagem natural e em sentenças KIF. O texto em linguagem natural é repassado a partes do sistema que têm um local para documentação que traduz as sentenças declarativas em axiomas.

Um terceiro exemplo de ferramenta para edição de ontologias é o *OntoEdit* desenvolvido pelo *Institut für Angewandte Informatik und Formale Beschreibungsverfahren* (AIFB) da Universidade de *Karlsruhe*. É disponibilizado nas versões *OntoEdit Free* que não possui suporte para inferências e o *OntoEdit Professional* que possui um editor gráfico de regras para editar axiomas F-Logic e o mecanismo de inferência *Ontobroker*. Utiliza como linguagem nativa a F-Logic, uma linguagem dedutiva de banco de dados orientado a objetos, que combina a semântica declarativa de bancos de dados com os recursos de modelagem do modelo de dados orientado a objetos. Importa e exporta para F-Logic, XML, RDF(S) e DAML+OIL (SURE *et al.*, 2003; KIFER *et al.*, 1995).

Em relação às linguagens, existem três tipos de linguagens utilizadas para a representação de ontologias: as tradicionais, as padrão *Web* e as *Web-based*. Entre as linguagens tradicionais cita-se Cycl, Ontolândia, F-Logic, CML, OCML, Loom e KIF. Entre as linguagens padrão *Web* a XML e RDF. E, entre as linguagens de ontologias *Web-based* cita-se OIL, DAML+OIL, SHOE, XOL e OWL (Almeida & Bax, 2003).

Entre as linguagens citadas a *Ontolândia*, RDF e OWL são recomendadas pelo *World Wide Web Consortium* (W3C). A Ontolândia é baseada em KIF (*Knowledge Interchange Format*) e é considerada a mais expressiva. Possibilita a representação de conceitos, taxonomias, relações n-árias, axiomas, instâncias e procedimentos. Não é fornecido suporte de raciocínio com a linguagem. A RDF (*Resource Description Framework*) foi projetada pelo W3C com o intuito de criar um modelo de dados simples, com semântica formal, com vocabulários baseados em URIs (*Uniform Resource Identifier*) e sintaxe baseada em XML (*Extensible Markup Language*). Trata-se de uma linguagem com bastante expressividade e possibilita a representação de conceitos, taxonomias, relações binárias, axiomas, instâncias e procedimentos. Adicionalmente estão sendo desenvolvidas máquinas de inferência para checagem de restrições. A OWL (*Ontology Web Language*) foi baseada nas linguagens OIL e DAML+OIL. É uma linguagem projetada para uso de aplicações que precisam processar informações da *Web* com interpretação por máquinas e fornece vocabulário adicional com semântica formal (MORAIS & AMBRÓSIO, 2007).

Uma linguagem que vem sendo utilizada recentemente para a construção e de ontologias é a UML (*Unified Modeling Language*) ou Linguagem de Modelagem

Unificada. A UML é uma linguagem estabelecida como padrão pelo *Object Management Group* (OMG) para a elaboração e estruturação lógica de sistemas de informação e é a mesma utilizada como base para o modelo OMT-G que representa o modelo conceitual brasileiro, a ET-EDGV. Estudos têm sido realizados (VAN NGUYEN *et al.*, 2016; MURPHY, 2009; FAHAD, 2008; CHUJAI *et al.*, 2008; GAEVI *et al.*, 2004) no sentido de transformar os componentes de modelos conceituais de banco de dados (entidades, atributos e relacionamentos) em linguagem UML para ontologias equivalentes em linguagem OWL (MKHININI *et al.*, 2020; VO & HOANG, 2019).

Cita-se ainda, a linguagem padrão *SPARQL Protocol and RDF Query Language* (SPARQL) definida pelo *World Wide Web Consortium* (W3C) para a realização de consultas à dados em formato RDF. Os resultados das consultas SPARQL podem ser conjuntos de dados ou triplas em RDF. Muitas pesquisas têm sido realizadas com o intuito de traduzir consultas em linguagem natural para consultas em formato SPARQL que retornem os resultados de acordo com raciocínio humano (POTONIEC *et al.*, 2020; RODRÍGUEZ DÍAZ *et al.*, 2019).

A escolha da linguagem depende do tipo de ontologia a ser especificada, dado que as linguagens utilizam suas componentes e raciocínio de diferentes maneiras. Do mesmo modo como ocorre entre os sistemas de geoinformação, há problemas de interoperabilidade entre os sistemas e linguagens utilizados para a geração de ontologias. O padrão XML foi desenvolvido com o objetivo de resolver esse problema e tem sido amplamente utilizado para a representação de ontologias (MORAIS & AMBRÓSIO, 2007; LIMA *et al.*, 2002).

Embora a linguagem XML seja menos expressiva que a OWL, a XML é mais popular na internet tanto para geração de arquivos de configuração, quanto para intercâmbio de dados entre aplicações ou para a estruturação e armazenamento de dados. No contexto geoespacial, o XML é favorável porque é um padrão adequado para descrever estruturas de dados hierárquicas e complexas, como é caso dos dados geoespaciais. Com a constante evolução da Web e o desenvolvimento da Web semântica, a XML tende a se tornar uma linguagem universal para a representação de dados (SOUSA *et al.*, MORAIS & AMBRÓSIO, 2007; LIMA *et al.*, 2002).

Acrescenta-se ainda a linguagem GML (*Geography Markup Language*) derivada da XML e específica para dados geoespaciais. De acordo com o *Open Geospatial*



*Consortium*, a GML serve como uma linguagem de modelagem para sistemas de geoinformação, assim como um formato de intercâmbio aberto para transações geoespaciais na *internet*. Como na maioria das gramáticas baseadas em XML, ela consiste de duas partes, o esquema que descreve o documento e o documento de instância que contém os dados. Desse modo, um documento GML é descrito usando um esquema GML (OGC, 2020).

Uma vantagem adicional da GML é a possibilidade de utilização de *tags* para expressar o significado dos dados descritos, agregando valor semântico a esses dados. Entretanto, em uma troca de dados entre diferentes usuários seria necessário fazer uma compatibilização semântica entre os modelos conceituais. Sob esse aspecto, Lima *et al.* (2002), citam o seguinte exemplo:

“Dois usuários de domínios diferentes representam uma determinada entidade, pela GML, como <rio> e <curso\_de\_agua>. Em uma troca de dados entre os usuários, os Esquemas também devem ser compartilhados, pois só assim uma aplicação poderá saber que <rio> ou <curso\_de\_agua> são da classe <\_Feature> definida pelo Esquema Feature.xsd da GML, e então usá-los adequadamente. Desta forma o problema de acesso aos dados é resolvido. Mas como saber que <rio> é <curso\_de\_agua> e vice-versa? O aspecto semântico não é considerado de forma efetiva a promover a interoperabilidade. Para amenizar este problema, pode-se acrescentar *tags* que descrevem as entidades e suas relações, ou que identifiquem sinônimos. Mas dentro de um ambiente heterogêneo, o ideal seria a criação de uma forma unificada ou padronizada de realizar isto, definida também nas especificações do *OpenGIS*”.

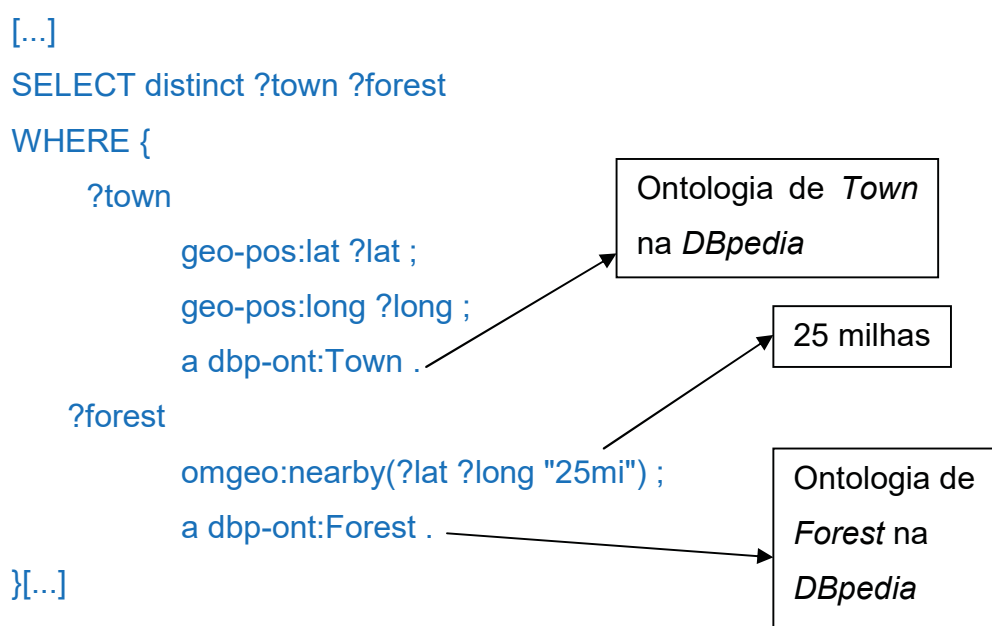
Esse exemplo demonstra a relevância do alinhamento semântico proposto nesta pesquisa de doutorado, porque embora existam recursos computacionais para descrever os dados semanticamente como as tags e o compartilhamento dos esquemas no caso da GML, a máquina, como usuária dessas informações, não tem capacidade para inferir sozinha que os significados das feições das diferentes fontes são sinônimos ou equivalentes. A base de conhecimento (knowledge base) e a ontologia, resultantes do alinhamento, servem para gerar novas inferências e ferramentas no contexto da inteligência artificial de modo a automatizar processos conforme sugerido por Lima et al. (2002) no exemplo supracitado.

Ainda sob esse aspecto, Gruber (1993b) cita um exemplo pertinente. Trata-se da reutilização de um sistema baseado em conhecimento para planejamento. O sistema faz uso de descrições de objetos, eventos, recursos e restrições para gerar



planos que atribuem recursos e horários a eventos e objetos. Embora seja possível utilizar algoritmos de planejamento geral, como todos os sistemas baseados em conhecimento, o planejamento em questão demanda uma base de conhecimento específica, nesse caso, geoespacial (a chamada "teoria de domínio" ou "conhecimento de base") para realizar as tarefas. Esta base de conhecimento deve conter conhecimentos genéricos para a tarefa de planejamento e outros que descrevem as situações de domínio. Contudo, para a eficácia na realização das tarefas precisa haver ainda, consenso na utilização dos termos entre as bases de conhecimento, necessitando assim de uma compatibilização semântica entre eles.

Um último exemplo, descreve resumidamente um caso real. Esse exemplo traz algumas especificidades da área de geoinformação e demonstra também que embora as ontologias sejam uma perspectiva relevante para a representação do conhecimento, muitas vezes são ainda necessárias estratégias para fazer acordos na troca de informações entre elas. Supõe-se que um usuário esteja estudando o papel da indústria florestal na economia rural. Supondo-se também que os termos floresta, cidade (pequena), fazenda e zona rural (*forest, town, farm and countryside*) sejam utilizados sem ter os seus significados pretendidos explicitados. Supõe-se ainda, que o usuário deseje fazer uma busca por “cidades próximas (*nearby*) de florestas” como na consulta em linguagem *SPARQL* na ontologia de Floresta da DBpedia, descrita abaixo e que as cidades recuperadas serão utilizadas para a análise (JANOWICZ *et al.*, 2013). A DBpedia é um esforço colaborativo para extrair conteúdo estruturado das informações criadas em vários projetos da *Wikimedia*.



De acordo com os autores, Janowicz *et al.* (2013) essa busca falharia pois a classe floresta (*Forest*) e outras similares não estão definidas na *DBpedia*. De qualquer modo, não importa o que a busca iria retornar, os resultados seriam enganosos. Entre todas essas pequenas cidades, o conjunto de dados também irá conter Los Angeles, Stuttgart e outras grandes metrópoles. Stuttgart, por exemplo, é descrita como a sexta maior cidade da Alemanha na *DBpedia*, mas classificada como uma cidade pequena nos dados interligados (*linked data*) também da *DBpedia*. Outro motivo para esse resultado equivocado é porque as classes cidade (*city*) e cidade pequena (*town*) foram definidas como equivalentes pela lei da Califórnia. A situação para o termo floresta (*forest*), reflorestamento (*afforestation*) e termos semelhantes é ainda mais complicada, com mais de novecentas definições, geralmente juridicamente vinculadas.

Para definir a relação espacial “próximo, perto de” (*nearby*) entre as cidades e as florestas nessa consulta foi utilizado 25 milhas (aproximadamente 40km). A exemplo do que ocorre com outros termos espaciais, a semântica do termo “próximo” também depende do contexto. Além disso, a maior parte dos dados interligados atualmente representa as feições geográficas pelos centróides, ao invés dos polígonos. Sendo assim, mesmo que um SIG represente uma cidade e uma floresta com dois polígonos adjacentes, ou seja, mais do que perto, seus centróides vão estar distantes de quilômetros (JANOWICZ *et al.*, 2013).

Conforme explicado anteriormente, na prática uma ontologia pode ser representada através de linguagem natural. No caso desta pesquisa, a ontologia foi representada através de linguagem natural e gráfica. As representações em linguagem natural e gráfica constam de toda a tese, com destaque para a seção de metodologia com os respectivos resultados e dos apêndices. Toda a documentação registrada através dos meios citados permitirá que futuramente a ontologia seja representada através de uma linguagem computacional. A etapa sobre a documentação apresenta mais detalhes sobre esta questão. Esta etapa não tem um resultado direto na seção de Resultados e Discussões.

### 3.4 INTEGRAÇÃO COM ONTOLOGIAS EXISTENTES

A quarta etapa é a de uma eventual ou intencional integração com outras ontologias. O objetivo da integração pode ser aproveitar conceituações preexistentes (FALBO, 1998), prover uma representação comum para a troca de informações ou para uma aplicação específica (FELICÍCIMO, 2004). Este é o caso da presente pesquisa, cujo resultado é uma ontologia que integra as ontologias do mapeamento topográfico oficial brasileiro (ET-EDGV 3.0) e do mapeamento colaborativo do OSM.

O objetivo da integração foi, com base nas conceituações preexistentes, prover uma representação comum para a troca de informações e para o desenvolvimento de novas ferramentas de inteligência artificial e sistemas baseados em conhecimento, bem como de aplicações que possibilitem a integração entre os bancos de dados dos mapeamentos utilizados e desses com outros bancos de dados para atualização e completude.

Conforme tratado na seção Interoperabilidade de Ontologias, a integração entre as ontologias dos mapeamentos, através dos seus modelos conceituais, pode ser classificada como um alinhamento de ontologias, dado que, resulta em uma estrutura formal com expressões que conectam os termos entre as ontologias mapeadas. Os resultados desta etapa são os mesmos da etapa de Captura da Ontologia.

### 3.5 AVALIAÇÃO DA QUALIDADE DA ONTOLOGIA

A quinta etapa é a de avaliação da qualidade da ontologia, a qual tem como intuito aferir se os resultados da ontologia gerada atendem ao propósito e aos requisitos especificados na primeira etapa. A avaliação da ontologia pode ser realizada de modo iterativo, concomitantemente as etapas de captura e formalização, estabelecendo-se critérios predefinidos (GUIZZARDI, 2005).

A avaliação da ontologia gerada nesta pesquisa foi realizada através de três mecanismos. O primeiro deles foi a aplicação de testes de alinhamento semântico entre os mapeamentos sobre distintas áreas de estudo utilizando um *plugin*. O segundo mecanismo de avaliação da ontologia foi a classificação quanto a o nível de

difficuldade do alinhamento semântico. E, o terceiro foi a aplicação de um *checklist* para detecção de inconsistências na ontologia, sugerido na literatura pesquisada.

Os testes de alinhamento semântico foram realizados através da aplicação de um *plugin*. Este *plugin* automatiza a integração do banco de dados do mapeamento colaborativo da plataforma OSM ao banco de dados do mapeamento topográfico oficial do Brasil de acordo com a Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV 3.0), com base no alinhamento semântico produzido nesta pesquisa. A automatização efetuada pelo *plugin* pressupõe que o banco de dados oficial já está em conformidade com a ET-EDGV 3.0.

O *plugin* foi desenvolvido no contexto dessa pesquisa com orientação da Prof<sup>a</sup> Dr<sup>a</sup> Silvana Philippi Camboim e co-orientação da autora desta tese como Trabalho de Conclusão de Curso de Engenharia Cartográfica e Agrimensura e como Estágio Supervisionado Obrigatório pelo aluno Vitor de Araújo Silva. Para este produto, resultante da aplicação do alinhamento semântico produzido nessa pesquisa, foi gerado também um artigo (SILVA *et al.*, 2020). O artigo descreve em detalhes o desenvolvimento e funcionamento do *plugin* para automatização do processo de integração dos mapeamentos e encontra-se no prelo.

Os testes de alinhamento atendem ao sexto objetivo específico da tese e foram realizados no ambiente do *software* QGIS. Os testes foram realizados nos dias 10 e 11 de Fevereiro de 2020.

No primeiro teste, foram utilizados dados do Instituto de Pesquisa e Planejamento Urbano de Curitiba (IPPUC) do município de Curitiba - PR (IPPUC, 2015). A base cartográfica da cidade é resultado de diversos mapeamentos, com vôos realizados entre 1972 e 2000, nas escalas 1:5.000, 1:2000 e 1:1.000<sup>1</sup>.

Foram selecionados os dados das feições (da coluna do banco de dados) do Sistema Viário Classificado, os quais foram compatibilizados com a ET-EDGV 3.0, pois os dados do IPPUC ainda não estão de acordo com a especificação.

Os dados do OSM foram importados para o QGIS utilizando o *plugin* QuickOSM. Em seguida, os dados do OSM foram compatibilizados de acordo com a ET-EDGV 3.0, utilizando o *plugin* desenvolvido no contexto da pesquisa. Foram

---

<sup>1</sup> IPPUC. Instituto de Pesquisa e Planejamento Urbano de Curitiba - Setor de Geoprocessamento. **Informações sobre mapas** [Mensagem pessoal]. Mensagem recebida por: <adri.alexandria@gmail.com>. 8 jan 2018.

utilizadas as etiquetas formadas pela chave *highway* (vias terrestres) para as classes *via\_deslocamento* e *trecho\_rodoviario*. Posteriormente foi aplicado um *buffer* de 15 metros nas vias, para uma escala estimada entre 1:20.000 e 1:30.000 do OSM no Brasil (ELIAS, 2018). Trata-se de um valor médio aproximado, pois a literatura mostra que diferentes variáveis podem alterar a acurácia posicional do OSM, entre elas a área de estudo, a época de edição dos dados e a dispersão da amostra.

Por fim, para cada segmento do *OpenStreetMap* foi aplicada a operação *Spatial Join* ou união espacial de atributos, para identificar qual era a classificação oficial daquele segmento e desse modo, poder efetuar a compararação com a classificação do *plugin*.

O segundo teste foi realizado para o município de Campo dos Goytacazes - RJ, utilizando a Base Contínua do Estado do Rio de Janeiro, produzida pelo IBGE na escala 1:25.000 (IBGE, 2018). Os dados dessa base já estão de acordo com a classificação da ET-EDGV 3.0. As demais etapas para a geração deste teste foram as mesmas realizadas para o primeiro teste.

Os resultados da aplicação dos testes foram compilados em duas tabelas com as porcentagens de feições classificadas correta e incorretamente. Foram também apresentados mapas esquemáticos das feições para o segundo teste.

O segundo mecanismo para a etapa de avaliação da qualidade da ontologia, foi a classificação quanto ao nível de dificuldade do alinhamento semântico entre as duas fontes. A classificação teve como base as similaridades e discrepâncias nos textos das descrições semânticas. Como resultado foi gerado uma tabela com três níveis de dificuldade (baixa, intermediária e alta), subdivididas pelas categorias da ET-EDGV 3.0 contendo as chaves do OSM alinhadas dentro de cada uma das categorias. Essa classificação é importante porque pode ajudar a determinar quais as etiquetas e/ou classes, atributos e valores de domínio podem levar a erros no alinhamento semântico.

O terceiro mecanismo utilizado para avaliação da ontologia foi a detecção de inconsistências no alinhamento entre as ontologias, conforme sugerido por Felicísimo (2004). De acordo com a autora, esse *checklist* considera alguns aspectos para minimizar os possíveis problemas nos processos de compatibilização entre as ontologias, como a presença de inconsistências nas comparações dos termos das ontologias, a falta de completude no alinhamento de todos os termos e, risco de inviabilidade da estratégia frente às inconsistências existentes.

Foi feita uma avaliação geral de todas as classes, dado que a compatibilização considerou previamente as diferenças entre os modelos conceituais antes de realizar o alinhamento semântico. Como resultado foi apresentado uma tabela listando as possíveis inconsistências e quais foram detectadas, seguidas das respectivas análises.

### 3.6 DOCUMENTAÇÃO DA ONTOLOGIA

A sexta e última etapa se refere à documentação de todas as etapas de geração da ontologia e, a exemplo da etapa de avaliação, deve ser feita paralelamente a cada uma das fases anteriores. A documentação deve incluir o propósito da ontologia, o levantamento dos requisitos e cenários de motivação, as descrições dos conceitos, a ontologia formal e os critérios de avaliação (GUIZZARDI, 2005).

De acordo com MORAIS & AMBRÓSIO (2007) deve ser observado o princípio do vocabulário mínimo e o princípio da auto-referência na documentação da ontologia e se possível discriminados em um Dicionário de Termos. O princípio do vocabulário mínimo trata dos termos utilizados na conceituação do domínio e o princípio da auto-referência orienta que um termo no Dicionário de Termos deve ser feito utilizando outros termos já definidos anteriormente. Os autores dizem ainda, que uma abordagem potencial para a documentação de ontologias é o uso de hipertextos: “esta tecnologia mostra-se adequada, já que torna natural a definição de novos termos a partir de outros mais primitivos, e permite a navegação entre definições, exemplos e a formalização, incluindo os axiomas”.

Conforme explicado na seção de formalização da ontologia, toda a documentação da ontologia de alinhamento entre as ontologias do mapeamento colaborativo da plataforma *OpenStreetMap* e do mapeamento topográfico oficial brasileiro (ET-EDGV 3.0) está documentado ao longo desta tese e dos apêndices em linguagem natural e gráfica.

O quinto objetivo específico da pesquisa é gerar uma base de conhecimento e uma ontologia da compatibilização entre os modelos conceituais para ser compartilhada, reutilizada e atualizada colaborativamente. Toda a documentação supracitada se constitui em uma base de conhecimento.

Toda essa documentação foi disponibilizada no repositório *GitHub* com o intuito de promover a sua discussão, o seu compartilhamento, atualização e manutenção de modo colaborativo.

O *GitHub* é um serviço *web* que hospeda gratuitamente projetos, *frameworks* e bibliotecas cujo desenvolvimento é *open source* e que possibilita que os trabalhos nele hospedados sejam desenvolvidos de modo colaborativo. *Git* é um sistema distribuído de controle de versão de arquivos (GITHUB, 2020).

Além disso, os resultados também foram disponibilizados na Base de Dados Científicos (BDC) da Universidade Federal do Paraná (UFPR, 2020), uma iniciativa que tem o intuito de reunir e disponibilizar dados científicos utilizados nas pesquisas publicadas pela comunidade acadêmica.

## 4. RESULTADOS E DISCUSSÕES

Nas próximas seções são apresentados os resultados referentes a cada uma das etapas descritas na metodologia, bem como, as discussões a respeito desses resultados.

### 4.1 IDENTIFICAÇÃO DE PROPÓSITO E ESPECIFICAÇÃO DE REQUISITOS

Nesta etapa foram identificados os usos e propósitos da ontologia, seus aspectos relevantes, potenciais usuários e o contexto que promoveu a sua criação. O resultado da etapa de identificação de propósito e especificação de requisitos é apresentado na tabela 1 e, descrito e analisado nos próximos parágrafos.

De acordo com as classificações descritas na seção 2.1.2.3. Tipos de Ontologias, o produto dessa pesquisa é uma ontologia **semi-informal**, em relação ao grau de formalismo, pois é expressa em uma forma restrita (no sentido de utilizar o vocabulário de domínio específico da geoinformação) e estruturada de linguagem natural.



TABELA 1 – IDENTIFICAÇÃO DE PROPÓSITO E ESPECIFICAÇÃO DE REQUISITOS

IDENTIFICAÇÃO DE PROPÓSITO E ESPECIFICAÇÃO DE REQUISITOS	
<b>TIPO DE ONTOLOGIA</b>	Semi-informal.
<b>APLICAÇÃO</b>	De acesso comum à informação e; de aplicação.
<b>CONTEÚDO</b>	De informação; de aplicação e; de especificação.
<b>FUNÇÃO</b>	De domínio; de tarefas e; de aplicação.
<b>CONTEXTO QUE PROMOVEU A CRIAÇÃO DA ONTOLOGIA</b>	Utilização do mapeamento colaborativo da plataforma <i>OpenStreetMap</i> como uma alternativa de baixo custo para a produção, atualização e agregação de dados e atributos ao mapeamento topográfico oficial no Brasil.
<b>PROPÓSITO</b>	Alinhar semanticamente as ontologias do <i>OpenStreetMap</i> e da Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV 3.0) brasileira, bem como, produzir uma base de conhecimento estruturada de todo o seu processo de geração.
<b>USOS PRETENDIDOS</b>	<ul style="list-style-type: none"> <li>- Promoção da interoperabilidade entre os dados dos mapeamentos utilizados e desses com outros mapeamentos e dados geoespaciais;</li> <li>- Documentação da estrutura de conhecimento encapsulada nos modelos conceituais para a geração de novas ferramentas de automatização envolvendo a integração entre mapeamentos e a;</li> <li>- Geração de ferramentas de inteligência artificial em sistemas baseados em conhecimento como os de processamento de linguagem natural (PLN), mineração de dados (<i>data mining</i>) e aprendizado de máquina (<i>machine learning</i>).</li> <li>- Formalização da ontologia para geração de recursos para exploração e busca de dados na <i>internet</i>.</li> </ul>
<b>POTENCIAIS USUÁRIOS</b>	<ul style="list-style-type: none"> <li>- Instituições produtoras de mapeamento oficial em diversas escalas, que incluem as entidades dos governos estaduais e federal, bem como, gestores e técnicos municipais das 5.560 prefeituras brasileiras;</li> <li>- Empresas;</li> <li>- Os mais de cinco milhões de colaboradores globais do <i>OpenStreetMap</i>;</li> <li>- A comunidade científica;</li> <li>- Produtores de sistemas e tecnologia;</li> <li>- Todo e qualquer cidadão.</li> </ul>
<b>QUESTÕES QUE A ONTOLOGIA DEVE RESPONDER</b>	<ul style="list-style-type: none"> <li>- Qual a estrutura de conhecimento encapsulado nos modelos conceituais dos mapeamentos?</li> <li>- Qual o significado de cada um dos termos referentes aos conceitos, classes de feições, atributos, valores de domínio, <i>tags</i>, chaves, valores e relacionamentos em cada um dos modelos conceituais dos mapeamentos?</li> <li>- Qual o termo e o significado correspondente no outro mapeamento?</li> <li>- Quais as geometrias previstas para cada feição em cada um dos mapeamentos?</li> <li>- Quais os possíveis casos de correspondência geométrica entre os elementos em cada um dos modelos conceituais dos mapeamentos?</li> <li>- Quais as multiplicidades dos possíveis casos de correspondência semântica?</li> </ul>

FONTE: A autora (2020).



Quanto à aplicação, trata-se de uma ontologia **de acesso comum à informação**, já que torna o vocabulário das ontologias originais inteligível, conferindo conhecimento compartilhado dos termos. O termo inteligível é utilizado aqui no sentido que a ontologia explicita um vocabulário de domínio específico de modo a permitir que leigos e especialistas de outras áreas de domínio compreendam as informações referentes aos mapeamentos, por ela descrita. Ainda quanto à aplicação, a ontologia em questão também pode ser classificada como **de especificação**, pois alinha duas ontologias de domínio que podem ser usadas para documentação e manutenção no desenvolvimento de aplicações de *software* e sistemas baseados em conhecimento. Mais adiante na seção que apresenta os resultados da etapa de Avaliação da Qualidade da ontologia, é apresentado o *plugin* desenvolvido no contexto desta pesquisa, uma aplicação que tem como base a compatibilização semântica proposta nesta pesquisa e que automatiza o processo de integração entre os bancos de dados dos mapeamentos.

Em relação ao conteúdo, a ontologia pode ser classificada como **de informação**, pois especifica a estrutura de registros de um banco de dados; como também **de aplicação**, já que contém as definições necessárias para modelar conhecimento em uma aplicação, e ainda; **de domínio**, pois expressa conceitualizações específicas do mapeamento de referência e do mapeamento colaborativo do *OSM* no domínio geoespacial.

Quanto à função a ontologia produzida pode ser classificada de três modos. Primeiramente como **de domínio**, pois alinha conceitos e vocabulários relativos a um domínio específico. É importante destacar que esses conceitos e vocabulários já haviam sido previamente definidos e consistem dos modelos conceituais de cada um dos mapeamentos. Segundo, pode ser classificada como **de tarefas**, dado que expressa conceitualizações sobre a resolução de problemas utilizando o vocabulário relacionado a uma atividade ou tarefa genérica que é a integração entre os bancos de dados dos mapeamentos. E terceiro, pode ser classificada como **de aplicação**, pois descreve conceitos que dependem de domínio e de tarefa específicos para subsidiar a geração de novas ferramentas. Caso este, do *plugin* desenvolvido com base no alinhamento semântico e que automatiza a integração entre os bancos de dados dos mapeamentos.

O **contexto que promoveu a criação da ontologia** foi descrito em detalhes na introdução e justificativa da tese. Desse modo, sucintamente, o contexto é a

utilização do mapeamento colaborativo da plataforma *OpenStreetMap* como uma alternativa de baixo custo para a produção, atualização e agregação de dados e atributos ao mapeamento topográfico oficial no Brasil, bem como o contrário, a agregação de dados oficiais ao *OSM*.

A ontologia teve **como propósito** alinhar semanticamente as ontologias do *OpenStreetMap* e da Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais (ET-EDGV 3.0) brasileira, bem como, gerar uma base de conhecimento estruturada de todo o seu processo de produção, os quais permitem relacionar o conhecimento do usuário e produtor leigo do *OSM* ao conhecimento do usuário e do especialista do mapeamento topográfico brasileiro.

No que se refere aos **usos**, um primeiro uso é a promoção da interoperabilidade entre os dados dos mapeamentos utilizados e destes com outros mapeamentos e dados geoespaciais. Um segundo **uso** pretendido para a ontologia é documentar a estrutura de conhecimento encapsulado nos modelos conceituais para a geração de novas ferramentas de automatização envolvendo a integração entre mapeamentos. Um terceiro uso pretendido é a geração de ferramentas de inteligência artificial em sistemas baseados em conhecimento como os de processamento de linguagem natural (PLN), de técnicas de mineração de dados (*data mining*) e aprendizado de máquina (*machine learning*). E, um quarto uso pretendido é a futura formalização da ontologia, ou seja, representá-la através de uma linguagem computacional, a qual permitirá a geração de recursos para exploração e busca de dados na *internet*.

Quanto aos **potenciais usuários**, são eles as instituições produtoras de mapeamento oficial em diversas escalas, que incluem as entidades dos governos estaduais e federal, bem como, gestores e técnicos municipais das 5.560 prefeituras brasileiras; os mais de cinco milhões de colaboradores globais do mapeamento do *OpenStreetMap*; as empresas; a comunidade científica; os produtores de sistemas e tecnologia e; o usuário mais importante, todo e qualquer cidadão.

Em relação às **questões que a ontologia deve responder**, estão: Qual é a estrutura de conhecimento encapsulado nos modelos conceituais dos mapeamentos? Qual o significado de cada um dos termos referentes aos conceitos, classes, feições, atributos, valores de domínio, *tags*, chaves, valores e relacionamentos em cada um dos modelos conceituais dos mapeamentos? Qual o termo e o significado correspondente no outro mapeamento? Quais as geometrias

previstas para cada feição em cada um dos mapeamentos? Quais os possíveis casos de correspondência geométrica entre os elementos em cada um dos modelos conceituais dos mapeamentos? E por fim, quais as multiplicidades dos possíveis casos de correspondência semântica?

## 4.2 CAPTURA DA ONTOLOGIA

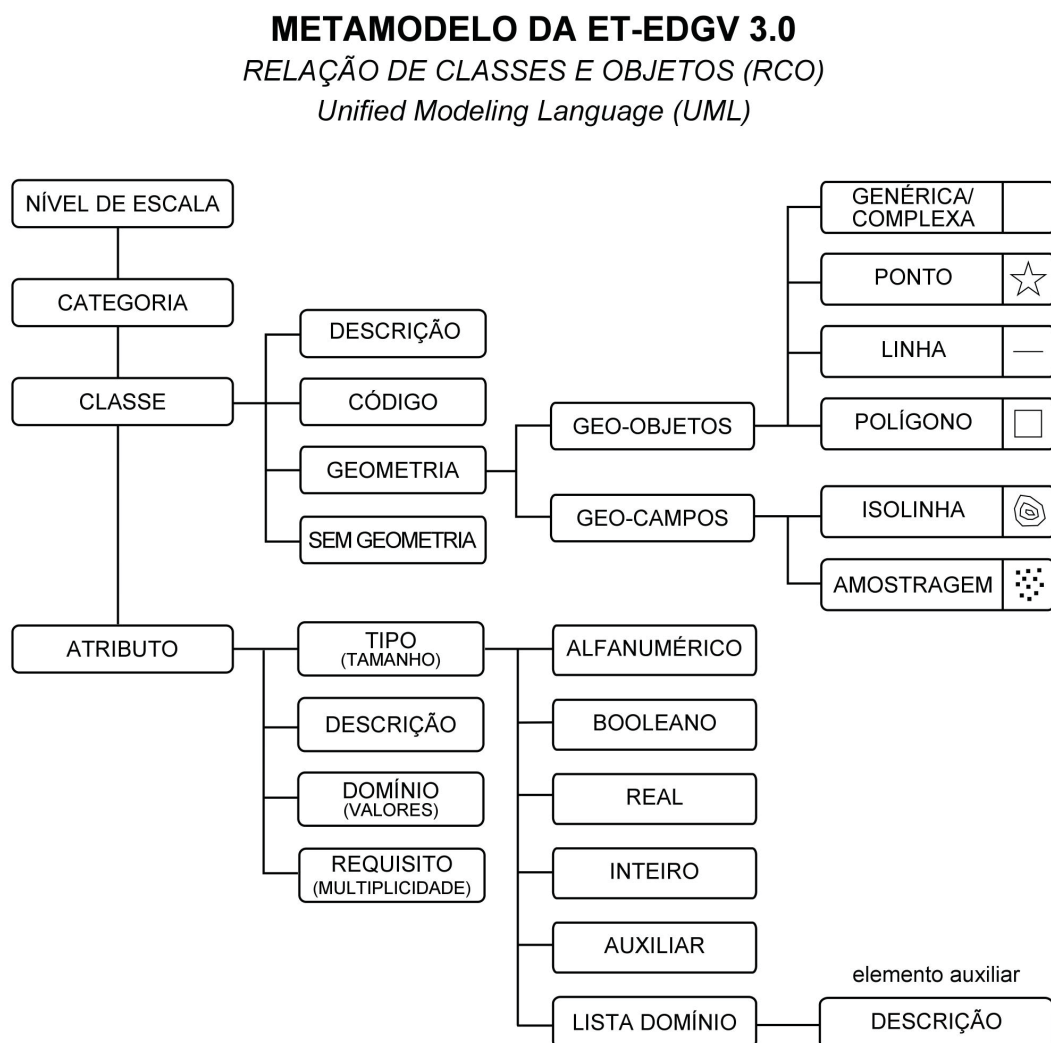
Nesta seção são apresentados os resultados e discussões referentes a segunda etapa metodológica Captura da ontologia. Os resultados estão divididos em três fases: (1) estrutura dos modelos conceituais; (2) seleção das chaves do *OSM* e categorias da ET-EDGV 3.0 e; (3) método proposto para o alinhamento semântico.

### 4.2.1 Estrutura dos modelos conceituais

Conforme descrito na metodologia, para alinhar a estrutura conceitual e semântica dos dois modelos conceituais foram produzidas representações gráficas. Essas representações são ilustradas pelas figuras 37, 38 e 39. Essas representações se constituem em resultados parciais da etapa de captura da ontologia e atendem ao objetivo específico número quatro, cujo intuito foi compor diagramas dos modelos conceituais, seus conceitos e relações, bem como do alinhamento semântico entre eles, e analisar os resultados do processo de integração. Estes três resultados estão disponíveis no Banco de Dados Científicos (BDC) da Universidade Federal do Paraná (UFPR) e tem um *Digital Object Identifier* (DOI) associado para facilitar o compartilhamento, a consulta e a citação.

Na figura 37 (<http://dx.doi.org/10.5380/bdc/49>) é apresentado um metamodelo da ET-EDGV 3.0 ilustrando sua estrutura conceitual e semântica. Na figura 37 tem-se o “nível de escala” que divide as categorias do mapeamento topográfico oficial brasileiro em escalas pequenas (MapTopoPE) e grandes (MapTopoGE). Conforme discutido anteriormente, essa divisão não deveria ocorrer, entretanto é a forma como esta versão da ET-EDGV é estruturada.

FIGURA 37 – METAMODELO CONCEITUAL E SEMÂNTICO DA ET-EDGV 3.0



FONTE: A autora (2020) (<http://dx.doi.org/10.5380/bdc/49>).

Dando continuidade a explicação da figura 37, na estrutura do modelo conceitual ET-EDGV 3.0 a segunda divisão hierárquica são as “categorias” de informação, as quais comportam conjuntos de classes ou grupos de feições divididos entre as pequenas e grandes escalas. As “categorias” de informações do MapTopoPE são: Energia e Comunicações, Estrutura Econômica, Hidrografia, Limites e Localidades, Pontos de Referência, Relevo, Saneamento Básico, Sistema de Transporte, Sistema de Transporte/ Subsistema Aeroportuário, Sistema de Transporte/ Subsistema Dutos, Sistema de Transporte/ Subsistema Ferroviário, Sistema de Transporte/ Subsistema Hidroviário, Sistema de Transporte/ Subsistema Rodoviário, Vegetação. As “categorias” do MapTopoGE são: Área Verde, Classes Base do Mapeamento Topográfico em Grandes Escalas, Classes Base do

Mapeamento Topográfico em Grandes Escalas/ Cultura e Lazer, Edificações, Classes Base do Mapeamento Topográfico em Grandes Escalas/ Estrutura de Mobilidade Urbana.

A terceira divisão hierárquica na estrutura do modelo conceitual ET-EDGV 3.0 são as “classes”. Como é possível observar na figura 37, para cada “classe”, há uma “descrição semântica”, um “código” que é um número que identifica a sua posição referente à categoria dentro do MapTopoPE ou do MapTopoGE e, a(s) “geometria(s)” que podem representar a feição. O tipo de representação da “geometria” pode ser de “geo-objetos” ou “geo-campos”. Para os “geo-objetos” são possíveis a geometria “genérica e/ou complexa”, “ponto”, “linha” ou “polígono”. Para os “geo-campos” são possíveis as geometrias “isolinha” ou “amostragem”. Os conceitos de todos esses elementos foram apresentados na seção sobre o modelo OMT-G na Fundamentação Teórica.

A quarta divisão hierárquica na estrutura da ET-EDGV 3.0 são os “atributos” para os quais existem as informações do “tipo (tamanho)”, a “descrição semântica” do atributo, os valores de “domínio” e o “requisito (multiplicidade)”. Para o “tipo (tamanho)” de atributo existem os valores: “alfanumérico”, “*booleano*”, “*real*”, “inteiro”, “auxiliar” e “lista de domínio”. Para as listas de domínio também existem as “descrições semânticas” como um elemento auxiliar ao entendimento do significado dos valores. Os conceitos desses elementos foram apresentados na seção sobre o Modelo Conceitual do Mapeamento Topográfico Oficial no Brasil. Todas essas informações constam das tabelas de cada uma das Classes dentro da ET-EDGV.

Na figura 38 (<http://dx.doi.org/10.5380/bdc/48>) é apresentado um diagrama ilustrando a estrutura do modelo conceitual e semântico do *OpenStreetMap*. Iniciando a leitura da figura pela parte superior central, tem-se inicialmente a escolha do “idioma”, dado que o OSM é uma plataforma global e a sua *wiki* tem as traduções de todo o modelo em diversos idiomas. Uma observação importante é que as chaves e valores são sempre no idioma de origem, ou seja, a língua inglesa, embora tenham as descrições semânticas traduzidas para outros idiomas. Algumas traduções no OSM são feitas através de tradutores virtuais e às vezes difíceis de compreender. Nesses casos, foi feita a leitura e a análise das descrições semânticas em inglês e, efetuada uma tradução livre com a respectiva indicação.



“nó”, “caminho”, “área” (que é um caminho fechado) e “relação”. Logo abaixo tem-se o “desenho” e a “foto” que são elementos auxiliares às descrições semânticas no modelo. O “desenho” se refere a um “exemplo da simbologia” que representa a feição no mapa do *OpenStreetMap*. A “foto” se refere à uma foto ou figura representando um “exemplo de vista da feição no terreno”.

Os diagramas das figuras 37 e 38 representam os modelos conceituais ET-EDGV 3.0 e do *OpenStreetMap* respectivamente, os quais permitiram compreender as suas estruturas semânticas e conceituais, bem como suas diferenças e semelhanças, com o intuito de sistematizar o processo de alinhamento semântico.

Devido às diferenças, o alinhamento entre os elementos não é direto e durante o processo foram constatadas diversas correspondências possíveis. A figura 39 (<http://dx.doi.org/10.5380/bdc/50>) representa essas possíveis correspondências entre os elementos estruturais e semânticos das ontologias da ET-EDGV e do OSM.

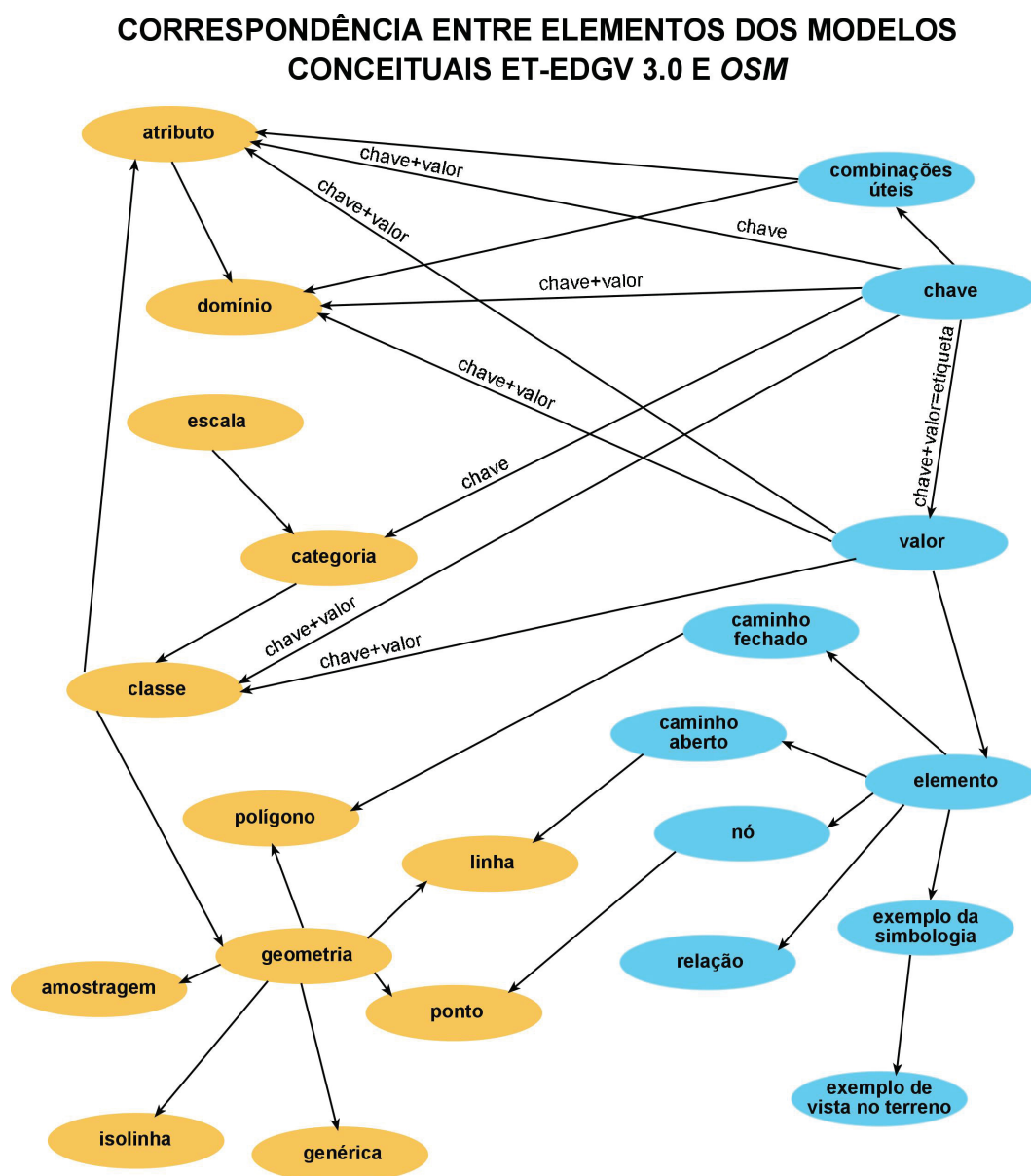
Na figura 39, em relação aos elementos semânticos, é possível observar que uma chave do OSM pode corresponder a uma categoria e/ou a um atributo na ET-EDGV. Do mesmo modo, uma chave+valor pode corresponder a um atributo, a um valor de domínio e/ou a uma classe. As combinações úteis podem corresponder a um atributo e/ou valor de domínio.

Em relação à geometria, o elemento de área (caminho fechado) corresponde ao elemento polígono na ET-EDGV, o elemento de caminho (aberto) corresponde ao elemento linha, o elemento de nó corresponde ao elemento ponto e o elemento de relação não tem correspondência na ET-EDGV 3.0. As geometrias de geo-campos, amostragem e isolinha da ET-EDGV 3.0 não têm correspondência no modelo do OSM.

Os elementos auxiliares às descrições semânticas “exemplo da simbologia” (desenho) e “exemplo de vista no terreno” (foto) do OSM não tem correspondência na ET-EDGV 3.0. O desenho se refere a um exemplo de como é simbolizada a feição no mapa e a foto é uma imagem fotográfica que fornece uma vista da feição no terreno. Esses são elementos que agregam um conteúdo informacional de grande valor ao modelo conceitual do OSM e que fazem falta na ET-EDGV 3.0, embora a simbologia faça parte do Manual Técnico T 34-700 – Convenções Cartográficas (DSG, 2002) e a Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais (ADGV 3.0) (DSG, 2018) tenha ilustrações das feições.



FIGURA 39 – CORRESPONDÊNCIA ENTRE OS ELEMENTOS CONCEITUAIS DA ET-EDGV 3.0 E DO OSM



FONTE: A autora (2020) (<http://dx.doi.org/10.5380/bdc/50>).

Na próxima seção são apresentados os resultados referentes à segunda fase da etapa de Captura da Ontologia.

#### 4.2.2 Seleção das chaves do OSM e categorias da ET-EDGV 3.0

O resultado da fase de seleção das chaves do OSM e categorias da ET-EDGV 3.0 encontra-se na tabela 2. A tabela 2 apresenta as chaves (*keys*) do *OpenStreetMap*, alinhadas dentro de cada categoria da ET-EDGV 3.0, assim como a porcentagem de classes e o número total de chaves do OSM alinhadas dentro de cada categoria da especificação brasileira. Ao todo, foram alinhadas 69 chaves do OSM e, 18 das 19 categorias da ET-EDGV 3.0.

TABELA 2 – CATEGORIAS E CHAVES ALINHADAS

CATEGORIAS E CHAVES ALINHADAS			
MAPEAMENTO TOPOGRÁFICO EM GRANDES ESCALAS			
Categoria ET-EDGV 3.0	Total de Classes	Chaves OSM	Chaves OSM
Área Verde	3 de 4 (75%)	<i>landuse, natural</i>	2
Classes Base do Mapeamento Topográfico em Grandes Escalas	13 de 29 (45%)	<i>landuse, amenity, animal, religion, power, building, industrial, man_made, bridge, bridge/structure, highway, key, maxspeed, tunnel, name, layer, covered, access, maxheight, maxwidth, maxweight, lit</i>	22
Cultura e Lazer	5 de 9 (56%)	<i>leisure, sport, highway, name, building, operator, operator/type, amenity, landuse, tourism</i>	10
Edificações	28 de 35 (80%)	<i>building, name, building/condition, abandoned, building/material, material, high, tourism, operator/type, addr, entrance, leisure, amenity, tourism, opening_hours, brewery, smoking, office, public_transport, highway, industrial, landuse, aeroway, natural/water, recycling_type, power, tourism, social_facility</i>	28
Estrutura de Mobilidade Urbana	8 de 9 (89%)	<i>highway, cycleway, aerialway, step_count, incline, handrail, surface, lit, ramp, building, public_transport</i>	11
MAPEAMENTO TOPOGRÁFICO EM PEQUENAS ESCALAS			
Categoria ET-EDGV 3.0	Total de Classes	Chaves OSM	Chaves OSM
Energia e Comunicações	11 de 16 (69%)	<i>building, power, man_made, man_made/tower_type</i>	4
Estrutura Econômica	2 de 3 (67%)	<i>industrial, building=material, landuse</i>	3
Hidrografia	8 de 20 (40%)	<i>landuse, natural, natural/water</i>	3
Limites e Localidades	4 de 25 (16%)	<i>leisure, boundary, landuse</i>	3
Relevo	7 de 18 (39%)	<i>natural</i>	1
Saneamento Básico	2 de 4 (50%)	<i>natural, natural/water, amenity</i>	3
Sistema de Transporte	10 de 16 (63%)	<i>public_transport, amenity, building, railway, building/condition, abandoned, surface, railway, name, operator, operator/type, aeroway, covered, access, capacity, fee, cyclestreets_id, maxstay, surveillance, bicycle_parking, bridge, man_made, bridge/structure, tunnel, highway</i>	25
Sistema de Transporte/ Aeroportuário	2 de 2 (100%)	<i>aeroway</i>	1
Sistema de Transporte/ Dutos	2 de 2 (100%)	<i>tunnel, man_made</i>	2
Sistema de Transporte/ Ferroviário	5 de 5 (100%)	<i>landuse, railway, public_transport, service, gauge, electrified, frequency, voltage, tracks, bridge, cutting, embankment, embedded_rails, tunnel, usage</i>	15
Sistema de Transporte/ Hidroviário	2 de 7 (29%)	<i>landuse, natural/water</i>	2
Sistema de Transporte/ Rodoviário	2 de 3 (67%)	<i>highway, surface, oneway, key, maxspeed</i>	5
Vegetação	7 de 13 (54%)	<i>natural, landuse</i>	2

FONTE: A autora (2020).

Dentre todas as categorias da ET-EDGV 3.0 a única categoria que não teve alinhamento de classe, atributo ou valor de domínio algum foi a de Pontos de Referência. Isso ocorreu por quatro motivos.

O primeiro é que a finalidade do mapeamento topográfico oficial é prover mapas de uso geral, nos quais, de acordo com Keates (1973), são representadas “todas as feições identificáveis da superfície da Terra, tanto naturais como artificiais, para as quais é possível estabelecer uma posição específica, expressa em relação à superfície topográfica”. O caráter oficial do mapeamento, entre outros aspectos, diz que seu propósito fundamental é representar as feições em suas posições, acurada e precisamente, dentro dos limites da escala. A categoria dos Pontos de Referência, de acordo com a definição da ET-EDGV 3.0 (DSG, 2017) “agrupa as feições que representam os elementos que servem como referência a medições em relação à superfície da Terra ou de fenômenos naturais”. A determinação dos Pontos de Referência é atribuída por lei às agências oficiais de mapeamento.

No OSM esses mesmos Pontos de Referência quando representados, não têm a finalidade intrínseca do mapeamento oficial, sendo mais informativa. A etiqueta *man\_made=survey\_point* é traduzida como “vértice geodésico ou marco geodésico” e, é descrita como “estruturas fixas, normalmente pilares, marcos ou chapas, determinadas com alta precisão e com altitude conhecida; são utilizados por exemplo na trigonometria e cartografia tradicional”. Existe também a chave *reference\_point* no *OpenStreetMap* mas ela é descrita como (WIKI OSM, 2020) “um ponto de referência cujo nome serve para orientação local”. São pontos que servem como uma alternativa a utilização de nomes de ruas para a orientação local em alguns países (grande parte da América Central, Nicarágua, Honduras, Guatemala, Costa Rica e Panamá) (WIKI OSM, 2020).

O segundo motivo pelo qual a categoria de Pontos de Referência foi a única que não teve alinhamento algum, é que não foram alinhadas todas as etiquetas referentes a algumas chaves, pois essas apenas eram complementares ou auxiliares a outras, como no caso das “combinações úteis” ou quando a descrição semântica do OSM sugere “ver também”. O terceiro é que foi alinhada uma amostra de 69 chaves e etiquetas que não compreende todo o universo semântico da ontologia do OSM. O quarto é que essa é uma categoria que tem menor potencial de integração, pois a determinação dos pontos de referência obedece à legislação e normas específicas para o seu levantamento e materialização.

Sob a luz dos estudos de organização do conhecimento (Rosch, 1973; Rosch *et al.*, 1976; Tversky & Hemenway, 1984) trata-se de uma categorização cujo grau de abstração é alto, portanto de nível superior. Não são feições facilmente reconhecíveis no mundo real e cuja inserção em uma categoria não ocorre de acordo com a sua semelhança com um protótipo da categoria.

Entre as categorias que tiveram baixa porcentagem de alinhamento estão: Limites e Localidades (16%), que é atribuição do Estado e dependente de legislação específica e; Vegetação (54%), Relevo (39%), Hidrografia (40%) e Sistema de Transporte/Hidroviário (29%), cujo levantamento depende de conhecimento técnico e/ou tecnologias específicas. Esses resultados vão ao encontro do estudo prévio de Bortolini *et al.* (2018). As demais categorias que também tiveram baixa porcentagem no alinhamento foram Saneamento Básico (50%), Cultura e Lazer (56%) e, Classes Base do Mapeamento Topográfico em Grandes Escalas (45%).

Em relação a esta última, possivelmente houve uma baixa porcentagem no alinhamento devido ao fato da estrutura da ET-EDGV 3.0 ser dividida nos grupos de Mapeamento Topográfico em Pequenas Escalas (MapTopoPE) e Mapeamento Topográfico em Grandes Escalas (MapTopoGE) conforme já discutido anteriormente. Além dessa subdivisão, esta categoria está dentro do MapTopoGE, a qual repete classes que fazem parte do MapTopoPE e agrega uma grande quantidade de classes, as quais não tiveram correspondentes no modelo do *OpenStreetMap* no alinhamento.

Sob o ponto de vista da estruturação do conhecimento, observa-se que, as categorias que tiveram baixa porcentagem no alinhamento, são categorias cujos termos são de nível superior de abstração na ET-EDGV 3.0 (hidrografia, relevo, vegetação e cultura e lazer, p.ex.) fruto de um raciocínio altamente abstrato na categorização.

No caso do OSM os termos das chaves alinhadas (para hidrografia, *natural* e *water*; para relevo, *natural*; para vegetação, *natural*; para cultura e lazer, *leisure*, *sport*, *amenity*, p. ex.) correspondem a “categorias naturais” (Rosch *et al.*, 1976), refletindo a estrutura do mundo real, mesmo para nomear as categorias em um nível hierárquico mais elevado. O uso desses termos é mais freqüente e fruto de um raciocínio não tão abstrato, característicos do nível básico de abstração. Uma categorização cuja validade de sugestão é mais alta, ou seja, com uma melhor

diagnose de associação à classe (ROSCH, 1973; ROSCH *et al.*, 1976; TVERSKY & HEMENWAY, 1984).

A maioria das categorias teve um bom resultado na quantidade de classes alinhadas. Dentre as dezoito categorias alinhadas treze tiveram cinquenta por cento ou mais classes com etiquetas do OSM que as corresponderam semanticamente, resultando em uma boa abrangência de classes dentro de cada categoria no alinhamento.

As categorias que tiveram alta porcentagem de alinhamento foram Sistema de Transporte/Aeroportuário (100%), Sistema de Transporte/Dutos (100%), Sistema de Transporte/Ferroviário (100%) Sistema de Transporte (63%), Energia e Comunicações (69%), Sistema de Transporte/ Rodoviário (67%), Estrutura Econômica (67%), todas do Mapeamento Topográfico em Pequenas Escalas. No caso do Mapeamento Topográfico em Grandes Escalas foram as categorias Estrutura de Mobilidade Urbana (89%) e Edificações (80%).

Todas essas categorias são compostas de feições artificiais, a maioria delas em áreas urbanas confirmando a abundante disponibilidade de dados do OSM principalmente nessas áreas e cuja classificação teve resultados próximos aos de algumas pesquisas na literatura (LUDWIG & ZIPF, 2019; ESTIMA & PAINHO, 2013). Área Verde (75%), também do Mapeamento Topográfico em Grandes Escalas teve uma boa porcentagem de alinhamento, com resultados discordantes dos obtidos por Ludwig & Zipf (2019) na classificação de áreas verdes urbanas na Alemanha e em Israel. Entretanto, existe concordância com esses autores no fato de que o contexto local e propósito de uso dos dados influenciam na análise e classificação final das feições.

Analisando-se os resultados das categorias que tiveram alta porcentagem de alinhamento sob o ponto de vista da estruturação do conhecimento, observou-se que pelo fato de as feições serem conhecidas e facilmente reconhecidas no mundo real (devido aos fatores supracitados), a categorização demanda um tipo de raciocínio menos abstrato, mais próximo do concreto. Esta é uma característica das categorias de nível básico, que carregam bastante informação para serem diferenciadas entre si, ou seja, que possuem alta validade de sugestão na diagnose de associação das feições em cada classe. Por exemplo, no contexto urbano esta característica é percebida na diferenciação entre uma área verde, e uma via, uma ferrovia, uma edificação, uma torre de energia e assim por diante. Os termos

utilizados na nomenclatura das classes e valores de domínio dentro das categorias no caso da ET-EDGV 3.0 e nas etiquetas do *OSM* são também utilizados com maior frequência na linguagem cotidiana (ROSCH *et al.*, 1976; TVERSKY & HEMENWAY, 1984).

Observou-se também que na ET-EDGV 3.0 e no modelo conceitual e semântico do *OpenStreetMap*, as estruturas de conhecimento, ou seja, a hierarquização das diversas categorias é feita por meio do mecanismo de classificação taxonômico. Percebeu-se que organização do grande número de categorias, classes e valores de domínio é feito de modo que seja possível realizar inferências através de ontologias e bancos de dados. Essa organização das informações demonstra que os conjuntos de categorias foram inicialmente analisados através de suas partes e outros atributos e posteriormente agrupados por tipos de acordo com a semelhança entre eles (TVERSKY & HEMENWAY, 1984).

#### 4.2.3 Método proposto para o alinhamento semântico

O método proposto para alinhamento semântico gerou três resultados. O primeiro resultado do método proposto para alinhamento semântico também foi disponibilizado no Banco de Dados Científicos (BDC) da UFPR e pode ser acessado através do *link* (<http://dx.doi.org/10.5380/bdc/53>). São tabelas que apresentam o resultado do alinhamento por divisão do mapeamento (MapTopGE e MapTopPE) e para cada categoria da especificação e chave principal do modelo conceitual do *OSM*. São duas tabelas no formato XLS com os dados e dezoito tabelas no formato PDF para visualização.

Para facilitar o fluxo de leitura, a figura 40 apresenta um exemplo dessas tabelas. Como é possível observar na figura 40 as tabelas estão estruturadas de acordo com a hierarquia da ET-EDGV 3.0 de modo que, têm as classes, atributos e valores de domínio da ET-EDGV 3.0 com suas respectivas descrições semânticas e as correspondentes etiquetas (chaves+valores) do modelo *OSM* com as respectivas descrições semânticas.

Conforme explicado anteriormente, devido às diferenças na estrutura das ontologias de cada um dos mapeamentos, não são todos os elementos que tem um correspondente no alinhamento. Em alguns casos é alinhada somente a classe, e/ou



somente o atributo, e/ou somente um ou mais valores de domínio. Em determinados casos existem elementos no *OSM* que podem funcionar como classes, atributos ou valores de domínio na ET-EDGV, mas que não existem na especificação brasileira, os quais são indicados na tabela. Quando o campo da tabela não está preenchido é porque não foi encontrada uma chave e/ou valor correspondente no *OSM* para aquele elemento da ET-EDGV.

FIGURA 40 – EXEMPLO DE TABELA DE RESULTADOS DO ALINHAMENTO SEMÂNTICO

descrições semânticas das principais  
chaves utilizadas no alinhamento  
dessa categoria

categoria da ET-EDGV 3.0  
e principais chaves do OSM alinhadas na tabela

ET-EDGV 3.0 – CLASSES BASE DO MAPEAMENTO TOPOGRÁFICO EM GRANDES ESCALAS												
OSM – CHAVES PRINCIPAIS LANDUSE E BUILDING												
ETIQUETA (TAG)					ATRIBUTOS							
CHAVE (KEY)	VALOR (VALUE)	DESCRIÇÃO OSM	CLASSE EDGV	DESCRIÇÃO EDGV	ATRIBUTO EDGV	DESCRIÇÃO EDGV	DOMÍNIO	DESCRIÇÃO DOMÍNIO EDGV	VALOR	CHAVE (KEY)	VALOR (VALUE)	DESCRIÇÃO TAG OSM
			Cemitério	Cemitério é um terreno ou recinto onde se guardam restos mortais.						landuse	cemetery	Cemitério: local para sepultar cadáveres. Sempre que possível adicionar a etiqueta de religião=religion* (ver os vários valores em Pt:Tag:amenity=place_of_worship). Em sepulturas individuais, fora de cemitérios usar amenity=grave_yard.
					nome	Indica o nome completo da instituição				name	=""	Nome: Utiliza-se name=* para atribuir um nome a elementos do mapa como por exemplo estradas, bares, estações ferroviárias, parques, edifícios e tudo o que tem um nome. Veja mais em Map:Feature#naming.
					geométria/geométrica	Indica que a geometria adquirida é aproximada em relação à escala prevista para o produto cartográfico						
					tipoCemitério	Indica o tipo de cemitério.						
							1) Desconhecido 2) Crematório	1) Valor desconhecido 2) "		amenity	crematorium	Crematório: lugar onde os corpos de pessoas falecidas são reduzidos a cinzas através de combustão. Muitas vezes estes espaços também prestam serviços memoriais.
										animal	crematorium	Crematório para animais. Ainda há uma discussão sobre quais etiquetas de animais / animais de estimação devem ser adicionadas para incinerações. Use com building = yes. Crematório para animais. Ainda há uma discussão sobre quais etiquetas de animais / animais de estimação devem ser adicionadas para incinerações. Use com building = yes. Especifique animais admitidos com tags como dog=yes/no, cat=yes/no, horse=yes/no, hamster=yes/no, possum=yes/no e assim por diante. Considere também adicionar name=* and operator*.
							3) Parque 4) Vertical	3) Local geralmente amplo e gramado, constituído por túmulos localizados no nível do solo. 4) Edifícios com o fim específico de comportar, geralmente em praças fechadas e justapostas, restos mortais.				

classe e descrição

atributo e descrição

valor de domínio e descrição

chave + valor correspondente e descrição

FONTE: A autora (2020) (<http://dx.doi.org/10.5380/bdc/53>).

No topo de cada tabela (figura 40) consta o nome da categoria da ET-EDGV 3.0 e logo abaixo as principais chaves do *OpenStreetMap* que foram alinhadas com esta categoria. Nas linhas subsequentes ao nome das chaves do *OSM*, se encontra uma breve descrição destas chaves. Esses topos das tabelas funcionam como um resumo do conteúdo de cada tabela. As tabelas 3,4 e 5 contêm esse resumo.



TABELA 3 – RESUMO DAS CATEGORIAS E CHAVES ALINHADAS

## PRINCIPAIS CHAVES DO OSM ALINHADAS COM AS CATEGORIAS DA ET-EDGV 3.0

MAPEAMENTO TOPOGRÁFICO EM GRANDES ESCALAS		
CATEGORIA ET-EDGV 3.0	CHAVES OSM	DESCRIÇÃO DAS CHAVES DO OSM
Cultura e Lazer	<i>leisure</i>	A chave lazer ( <i>leisure</i> ) é utilizada para locais onde as pessoas podem ir no seu tempo livre.
	<i>sport</i>	A chave esporte ( <i>sport</i> ) é utilizada para indicar a prática de uma modalidade de esporte/desporto específico.
Área Verde	<i>landuse</i>	A chave uso do solo ( <i>landuse</i> ) é utilizada para descrever o principal uso do solo.
	<i>natural</i>	A chave natural ( <i>natural</i> ) é utilizada para descrever características naturais, principalmente em termos de habitat e características geológicas. Inclui também características que tenham sido alteradas por humanos.
Classes Base do Mapeamento Topográfico em Grandes Escalas	<i>building</i>	A chave edifício ( <i>building</i> ) é utilizada para demarcar os contornos de um edifício. Ela serve para identificar o tipo da edificação, se habitacional, comercial, religiosa, etc.
	<i>landuse</i>	A chave uso do solo ( <i>landuse</i> ) é usada para descrever o uso principal do terreno pelo homem.
Edificações	<i>building</i>	A etiqueta edifício ( <i>building</i> ) é utilizada para demarcar os contornos de um edifício. Ela serve para identificar o tipo da edificação, se habitacional, comercial, religiosa, etc.
	<i>amenity</i>	A etiqueta infraestrutura ( <i>amenity</i> ) contém as etiquetas usadas para marcar infraestruturas usadas por visitantes e residentes de um local. Ela detalha o uso das edificações, como banco, hospital, escola, etc. Demais chaves: name, operator, addr, high, opening hours, tourism, leisure, brewery, smoking, office, industrial, public transport, landuse, highway, aeroway, natural, water, recycling type, power, social facility e religious.
Estrutura de Mobilidade Urbana	<i>highway</i>	A chave vias terrestres ( <i>highway</i> ) é a principal etiqueta usada para identificar qualquer tipo de estrada, rua ou caminho. O valor para a chave highway ajuda a indicar a importância da rodovia dentro da malha viária como um todo.
MAPEAMENTO TOPOGRÁFICO EM PEQUENAS ESCALAS		
Energia e Comunicações	<i>power</i>	A etiqueta energia ( <i>power</i> ) contém as etiquetas para Estas etiquetas são usadas para cartografar sistemas de geração e distribuição de energia.
	<i>man_made</i>	A etiqueta construção ( <i>man_made</i> ) contém as etiquetas que descrevem as estruturas construídas por humanos (artificiais).
	<i>building</i>	A etiqueta edifício ( <i>building</i> ) é utilizada para demarcar os contornos de um edifício. Ela serve para identificar o tipo da edificação, se habitacional, comercial, religiosa, etc.

FONTE: A autora (2020).

TABELA 4 – RESUMO DAS CATEGORIAS E CHAVES ALINHADAS (cont.)

**PRINCIPAIS CHAVES DO OSM ALINHADAS COM AS CATEGORIAS DA ET-EDGV 3.0**

<b>MAPEAMENTO TOPOGRÁFICO EM PEQUENAS ESCALAS</b>		
<b>CATEGORIA ET-EDGV 3.0</b>	<b>CHAVES OSM</b>	<b>DESCRIÇÃO DAS CHAVES DO OSM</b>
<b>Estrutura Econômica</b>	<i>landuse</i>	A etiqueta uso do solo ( <i>landuse</i> ) é usada para descrever o uso principal do terreno pelo homem.
	<i>industrial</i>	A etiqueta indústria ( <i>industrial</i> ) é utilizada para descrever o tipo de indústria.
<b>Hidrografia</b>	<i>landuse</i>	A chave uso do solo ( <i>landuse</i> ) é utilizada para descrever o principal uso do solo.
	<i>natural</i>	A chave natural ( <i>natural</i> ) é utilizada para descrever características naturais, principalmente em termos de habitat e características geológicas. Inclui também características que tenham sido alteradas por humanos.
	<i>water</i>	A chave água ( <i>water</i> ) é utilizada é usada para especificar que tipo de corpo d'água é marcado com a etiqueta natural=water, só devendo ser usada em conjunto com esta.
<b>Limites e Localidades</b>	<i>landuse</i>	A chave uso do solo ( <i>landuse</i> ) é utilizada para descrever o principal uso do solo.
	<i>boundary</i>	A chave limites ( <i>boundary</i> ) marca as fronteiras das áreas, principalmente políticas, mas também de outras áreas administrativas.
	<i>leisure</i>	A chave lazer ( <i>leisure</i> ) é usada em infraestruturas de lazer.
<b>Relevo</b>	<i>natural</i>	A chave natural ( <i>natural</i> ) é utilizada para descrever características naturais, principalmente em termos de habitat e características geológicas. Inclui também características que tenham sido alteradas por humanos.
<b>Saneamento Básico</b>	<i>amenity</i>	A etiqueta infraestrutura ( <i>amenity</i> ) contém as etiquetas usadas para marcar infraestruturas usadas por visitantes e residentes de um local. Ela detalha o uso das edificações, como banco, hospital, escola, etc.
	<i>natural</i>	A chave natural ( <i>natural</i> ) é utilizada para descrever características naturais, principalmente em termos de habitat e características geológicas. Inclui também características que tenham sido alteradas por humanos.
<b>Sistema de Transporte</b>	<i>highway</i>	A chave vias terrestres ( <i>highway</i> ) é a principal etiqueta usada para identificar qualquer tipo de estrada, rua ou caminho. O tipo (valor para a chave) highway ajuda a indicar a importância da rodovia dentro da malha viária como um todo.
<b>Sistema de Transporte/ Aeroportuário</b>	<i>aeroway</i>	A chave infraestrutura aeroportuária ( <i>aeroway</i> ) (tradução livre) é usada para etiquetar infraestrutura física destinada a dar suporte a aeronaves e viagens aéreas, em particular os elementos relacionados a aeroportos e heliportos, etc.

FONTE: A autora (2020).

TABELA 5 – RESUMO DAS CATEGORIAS E CHAVES ALINHADAS (cont.)

## PRINCIPAIS CHAVES DO OSM ALINHADAS COM AS CATEGORIAS DA ET-EDGV 3.0

MAPEAMENTO TOPOGRÁFICO EM PEQUENAS ESCALAS		
CATEGORIA ET-EDGV 3.0	CHAVES OSM	DESCRIÇÃO DAS CHAVES DO OSM
Sistema de Transporte/ Dutos	<i>tunnel</i>	A chave túnel ( <i>tunnel</i> ) é utilizada pra mapear canais que circulam no subsolo de estradas, linhas ferroviárias, etc.
Sistema de Transporte/ Ferroviário	<i>railway</i>	A chave transporte ferroviário ( <i>railway</i> ) é usada para marcar os caminhos para muitos tipos de estradas de ferro, incluindo transporte ferroviário, incluindo a luz, ferroviárias importantes, metrô, mon trilhos e elétricos.
	<i>landuse</i>	A chave uso do solo ( <i>landuse</i> ) é usada para descrever o uso principal do terreno pelo homem.
	<i>public_transport</i>	A chave transporte público ( <i>public_transport</i> ) é usada para identificar qualquer estação ferroviária / metrô, aeroporto, estação de ônibus, porto de passageiros, etc.
Sistema de Transporte/ Hidroviário	<i>natural</i>	A chave natural ( <i>natural</i> ) é utilizada para descrever características naturais, principalmente em termos de habitat e características geológicas. Inclui também características que tenham sido alteradas por humanos.
	<i>landuse</i>	A chave uso do solo ( <i>landuse</i> ) é usada para descrever o uso principal do terreno pelo homem.
Sistema de Transporte/ Rodoviário	<i>highway</i>	A chave vias terrestres ( <i>highway</i> ) é a principal etiqueta usada para identificar qualquer tipo de estrada, rua ou caminho. O tipo (valor para a chave) highway ajuda a indicar a importância da rodovia dentro da malha viária como um todo.
Vegetação	<i>natural</i>	A chave natural ( <i>natural</i> ) é utilizada para descrever características naturais, principalmente em termos de habitat e características geológicas. Inclui também características que tenham sido alteradas por humanos.
	<i>landuse</i>	A chave uso do solo ( <i>landuse</i> ) é usada para descrever o uso principal do terreno pelo homem.

FONTE: A autora (2020).

O segundo resultado do alinhamento entre as ontologias foi disponibilizado no BDC da UFPR através do link (<http://dx.doi.org/10.5380/bdc/52>). A figura 41 mostra, como exemplo, um trecho da tabela que contém estes resultados. É uma tabela que apresenta os possíveis Casos de Correspondência Semântica, com as suas respectivas multiplicidades e, os possíveis casos de correspondência geométrica entre as ontologias dos dois mapeamentos. Esta tabela é dividida pelas subdivisões do mapeamento (MapTopGE e MapTopPE) representadas na cor



amarela na figura 41, pelas categorias de informação (p. ex. Sistema de Transporte/Rodoviário) na cor laranja, e elementos semânticos e geométricos das duas ontologias (modelos conceituais) e os possíveis casos de correspondência de acordo com uma legenda, na cor verde. Essa tabela não contém as descrições semânticas dos elementos conceituais e permite uma visão geral e sistematizada do alinhamento entre esses elementos. Os próximos parágrafos explicam melhor o conteúdo da Tabela de Casos de Correspondência Semântica, que pode ser observado na figura 41.

FIGURA 41 – EXEMPLO DA TABELA DOS CASOS DE CORRESPONDÊNCIA SEMÂNTICA

1 MAPEAMENTO TOPOGRÁFICO EM PEQUENAS ESCALAS									
1.1 ENERGIA E COMUNICAÇÕES									
ET-EDGV 3.0				OSM			CORRESPONDÊNCIA		
Classe	Atributo	Domínio	G	Chave	Valor	G	Semântica	Multiplicidade	Geométrica
Casa_De_Forca	—	—	P	building	transformer_tower	a	1	1:1	7
Complexo_Gerador_Energia_Eletrica	—	—	C	power	compensator	n, a	1	1:1	6
Est_Gerad_Energia_Eletrica	—	—	A,L,P	power	plant	a, r	1	1:1	6
	tipoEstGerad	3) Solar	A,L,P	power	heliostat	n	3	1:1	8
Grupo_Transformadores	—	—	A,P	power	converter	n, a	1	1:m	5
Grupo_Transformadores	—	—	A,P	power	switch	n	1	1:m/m:1	8
Grupo_Transformadores	—	—	A,P	power	transformer	n	1	1:m/m:1	8
Subest_Transm_Distrib_Energia_Eletrica	—	—	C	power	substation	n, a	1	1:m	6
Subest_Transm_Distrib_Energia_Eletrica	—	—	C	power	connection	n	1	1:m	6
Subest_Transm_Distrib_Energia_Eletrica	—	—	C	power	busbar	c	1	1:m	6
Subest_Transm_Distrib_Energia_Eletrica	—	—	C	power	bay	c	1	1:m	6
Subest_Transm_Distrib_Energia_Eletrica	—	—	C	power	switchgear	a	1	1:m	6
Subest_Transm_Distrib_Energia_Eletrica	—	—	C	power	marker	n	1	1:m	6

FONTE: A autora (2020) (<http://dx.doi.org/10.5380/bdc/52>).

Os casos de correspondência semântica podem ser: entre uma classe da ET-EDGV e uma etiqueta do OSM (chave+valor), número 1 da legenda, na coluna de correspondência semântica na tabela; entre um atributo da ET-EDGV e uma etiqueta do OSM, número 2 na tabela; entre um valor de domínio e uma etiqueta do OSM, número 3 na tabela; e ainda um atributo e/ou valor de domínio que tem no OSM e não tem na ET-EDGV, número 4 na tabela.

A tabela casos de correspondência semântica, também apresenta as multiplicidades do alinhamento semântico, as quais têm as seguintes possibilidades:

correspondências exatas (1:1); uma chave+valor equivale a mais de um valor de domínio, atributo e/ou classe (m:1) e; mais de uma chave+valor equivale a um valor de domínio, atributo ou classe (1:m).

As geometrias das feições são descritas na coluna “G” e podem ser: A=área; P=ponto; L=linha; C=complexa; G=genérica; e conv=convencional, no caso da ET-EDGV 3.0. No caso do OSM as geometrias descritas na coluna “G” podem ser: a=área; c=caminho; n=nó e; r=relação.

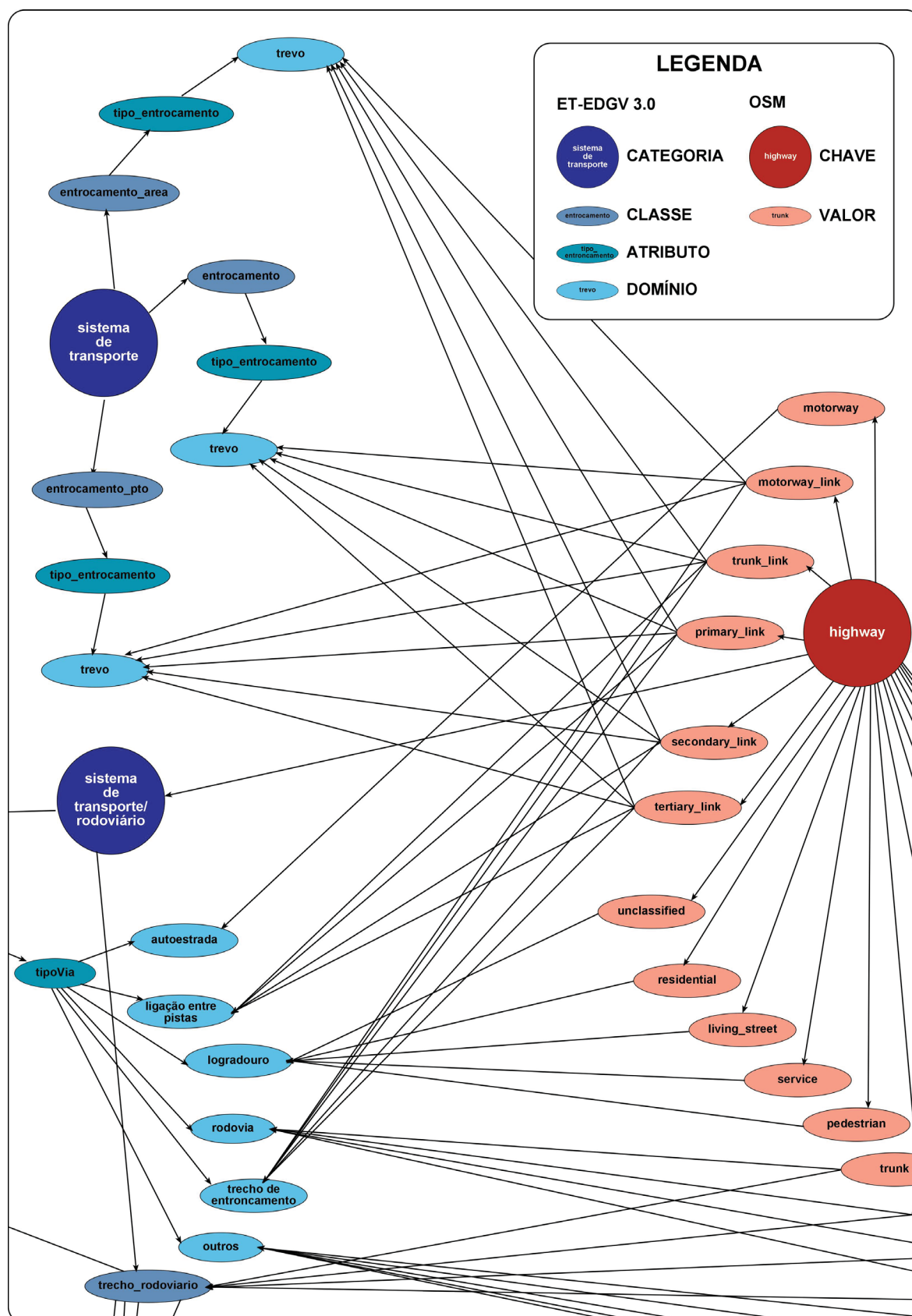
Em relação às correspondências geométricas, as feições podem ter: a mesma geometria, número 5 na coluna de correspondência geométrica; geometria complexa no caso da ET-EDGV ou, de modo similar ser uma relação no caso do OSM, número 6; geometrias diferentes com geometria mais complexa e/ou detalhada no OSM, número 7; geometrias diferentes com geometria mais complexa e/ou detalhada na ET-EDGV, número 8; ou sem geometria no caso da classe convencional na ET-EDGV número 0. No caso de geometria mais complexa e/ou detalhada tem-se, por exemplo, uma geometria de polígono na ET-EDGV e uma geometria de nó no OSM, que caso seja necessário ou desejável em uma aplicação, demandaria uma transformação.

O caso da classe convencional na ET-EDGV 3.0, de acordo com (DSG, 2017 p. 15) é uma classe que descreve um conjunto de objetos com propriedades, comportamento, relacionamentos e semântica semelhantes, as quais possuem alguma relação com os objetos espaciais, mas que não possuem propriedades geométricas. Essa situação ocorre apenas para duas classes na especificação brasileira. A primeira é a classe *Classif\_Econ\_Administ* que é (DSG, 2017 p. 112) “associada a cada edificação, sempre que for o caso, para classificar esta edificação quanto a sua administração, classe, divisão e grupo de atividade econômica”. A segunda é a classe *Equip\_Desenv\_Social* que tem a função de classificar uma instalação ou construção cujas atividades estão relacionadas ao atendimento de públicos e prestação de serviços das políticas de desenvolvimento social, com a sigla, código, localização e tipo de equipamento de desenvolvimento social (DSG, 2017 p. 126-127).

O terceiro resultado gerado pelo alinhamento semântico entre as ontologias do OSM e da ET-EDGV 3.0 foi também disponibilizado no Banco de Dados Científicos (BDC) da UFPR e pode ser acessado através do *link* (<http://dx.doi.org/10.5380/bdc/51>). Trata-se de um diagrama em forma de grafos

representando o alinhamento da chave *highway* do *OSM* com as categorias Sistema de Transporte e Sistema de Transporte/Rodoviário da ET-EDGV 3.0. Esse tipo de representação torna mais fácil a visualização da estrutura hierárquica e conceitual de cada modelo, o tamanho das classes e chaves, bem como as multiplicidades das correspondências semânticas entre eles no alinhamento. A figura 42 mostra uma parte desse grafo como exemplo.

FIGURA 42 – EXEMPLO DE PARTE DO GRAFO PARA O ALINHAMENTO ENTRE A CHAVE HIGHWAY DO OSM E A CATEGORIA SISTEMA DE TRANSPORTE DA ET-EDGV 3.0



FONTE: A autora (2020) (<http://dx.doi.org/10.5380/bdc/51>).



Na representação da figura 42, foram reservados os tons de vermelho para o modelo do OSM e os tons de azul para a ET-EDGV 3.0. A chave *highway* pode ser combinada com 21 valores, ou seja, resultando em 21 etiquetas com diferentes significados. As setas que saem de cada uma delas encontram uma ou mais classe, atributo e/ou valor de domínio correspondentes nas categorias Sistema de Transporte e Sistema de Transporte/Rodoviário da ET-EDGV 3.0. Por exemplo, na figura 42 é possível visualizar no grafo que a etiqueta *highway=tertiary\_link* teve uma multiplicidade m:1 (ver Tabela de Casos de Correspondência Semântica: <http://dx.doi.org/10.5380/bdc/52>), ou seja, uma chave+valor equivalente a mais de um domínio, atributo e/ou classe. Isso ocorreu porque a etiqueta *highway=tertiary\_link* foi alinhada semanticamente com o valor de domínio “trevo”, do atributo “tipo\_entroncamento”, da classe “entroncamento\_area”, da categoria Sistema de Transporte. A mesma etiqueta foi alinhada também com o valor de domínio “trevo”, do atributo “tipo\_entroncamento”, da classe “entroncamento\_pto”, da categoria Sistema de Transporte. Do mesmo modo, foi alinhada com o valor de domínio “trecho de entroncamento”, do atributo “tipo\_via”, da classe “via\_deslocamento” (que não pode ser vista no recorte apresentado na figura 42, mas pode ser vista no grafo), da categoria Sistema de Transporte/Rodoviário. E assim sucessivamente seguindo a seta que sai da etiqueta na direção de cada elemento das categorias Sistema de Transporte e Sistema de Transporte/Rodoviário da ET-EDGV 3.0.

#### 4.3 AVALIAÇÃO DA QUALIDADE DA ONTOLOGIA

Conforme apresentado na seção da Metodologia, a etapa de avaliação da qualidade da ontologia foi realizada através de três mecanismos. O primeiro deles foi a aplicação dos testes de alinhamento semântico sobre duas áreas de estudo utilizando o *plugin* desenvolvido no contexto desta pesquisa. A tabela 6 mostra os resultados do primeiro teste, realizado para o município de Curitiba - PR.

TABELA 6 – TESTE DE AVALIAÇÃO DO ALINHAMENTO CURITIBA

## TESTE DE AVALIAÇÃO DE QUALIDADE DO ALINHAMENTO - CURITIBA - PR

ETIQUETA (TAG)	Nº SEGMENTOS	CLASSIFICADO CORRETAMENTE	%	CLASSE ET-EDGV 3.0 ALINHADA	CLASSIFICADO INCORRETAMENTE trecho_arruamento trecho_rodoviário	%	CLASSIFICADO INCORRETAMENTE OUTROS		SEM CORRESPONDÊNCIA NA BASE OFICIAL	%
Residential	33875	33464	<b>98,79%</b>	trecho_arruamento	11	0,03%	16	0,05%	384	1,13%
Unclassified	600	549	<b>91,50%</b>	trecho_arruamento	10	1,67%	-	0%	41	6,83%
Trunk	805	201	24,97%	trecho_rodoviário	542	<b>67,33%</b>	-	0%	62	7,70%
Primary	2668	42	1,57%	trecho_rodoviário	2571	<b>96,36%</b>	-	0%	55	2,06%
Secondary	5710	24	0,42%	trecho_rodoviário	5582	<b>97,76%</b>	-	0%	104	1,82%
Tertiary	6635	11	0,17%	trecho_rodoviário	6541	<b>98,58%</b>	10	0,15%	73	1,10%
Motorway	237	199	<b>83,97%</b>	trecho_rodoviário	35	14,77%	-	0%	3	1,27%

FONTE: A autora (2020).

Na tabela 6 é possível observar que a melhor porcentagem resultante do alinhamento foi para a etiqueta *highway=residential*, que corresponde aos trechos de arruamento de vias residenciais, com aproximadamente 99% dos segmentos das vias classificados corretamente. O mesmo ocorreu para a etiqueta *highway=motorway*, que corresponde as autoestradas, com aproximadamente 84% dos segmentos classificados corretamente. E também para a etiqueta *highway=unclassified*, que no alinhamento corresponde a logradouro, com aproximadamente 92% dos segmentos classificados corretamente, embora a descrição semântica do OSM para essa etiqueta seja ambígua e tenha gerado dúvidas no alinhamento.

Sobre as classificações incorretas, as etiquetas *highway=trunk* (67%), *highway=primary* (96%), *highway=secondary* (98%) e *highway=tertiary* (99%), que foram alinhadas às rodovias e trechos rodoviários na ET-EDGV 3.0, há várias explicações para as discrepâncias. A primeira delas pode ter sido o fato de essas etiquetas terem descrições semânticas ambíguas e por este motivo terem sido alinhadas multiplamente com diferentes classes e valores de domínio. Outro fator é que possivelmente, além do alinhamento pode ser necessária uma análise de contexto na classificação para que um sistema consiga fazer a correspondência corretamente de forma automática. No caso de Curitiba, o teste foi feito para o município que é todo de área urbana, na qual diversas rodovias passaram para

gestão municipal, tendo classificações viárias próprias, situação que deve ter influenciado nos resultados. Outra possibilidade, é que pode ter havido também erros na classificação original dos dados pelos colaboradores.

Em face dos conceitos de organização do conhecimento (Rosch, 1973; Rosch *et al.*, 1976; MacEachren, 1995) observou-se que no caso da ET-EDGV 3.0 a categoria Sistema de Transporte/Rodoviário e no caso do *OpenStreetMap*, a chave *highway* ou “vias terrestres” que foi alinhada semanticamente com esta, são resultado de uma categorização de nível de abstração básico conforme analisado na seção 4.2.2., embora designem uma posição hierárquica mais elevada na organização taxonômica (categoria/ classe/ atributo/ valor de domínio na ET-EDGV e chave/ chave+valor no OSM).

Em relação às classes alinhadas, Via\_deslocamento e Trecho\_rodoviário, do modelo conceitual brasileiro, estas resultam em uma categorização que pode ser analisada de duas maneiras. A primeira análise poderia conferir a estas classes nível básico de abstração, pois tem formas similares, podem ser identificados por formas semelhantes e reconhecíveis, cujo nível de informatividade é maior e no qual, as descontinuidades entre as feições são mais salientes, por exemplo, vias de arruamento e rodovias diferenciadas de ferrovias. A segunda análise poderia conferir a estas classes, nível superior de abstração, pois são mais genéricas em relação aos valores de domínio a eles atribuídos (ROSCH *et al.*, 1976; TVERSKY & HEMENWAY, 1984).

Em relação aos valores das etiquetas (etiqueta=chave+valor) do OSM alinhadas com os valores de domínio correspondentes na ET-EDGV 3.0, ocorre uma variação na categorização. No caso dos valores que tiveram melhor porcentagem no alinhamento, por exemplo, *residential*, *unclassified* (logradouro na ET-EDGV 3.0) e, *motorway* (autoestrada), percebe-se que estes termos estão no nível de abstração básico, demandando um tipo de raciocínio intermediário entre o abstrato e o concreto. Neste nível as feições (uma rua e uma estrada) refletindo a estrutura do mundo real são facilmente diferenciadas. Já no caso dos valores que tiveram pior porcentagem alinhamento, por exemplo, *trunk*, *primary*, *secondary* e *tertiary* (alinhados com o valor rodovia e a classe trecho rodoviário na ET-EDGV 3.0) observa-se que, embora estejam no mesmo nível hierárquico da taxonomia da ET-EDGV 3.0, envolvem uma segunda hierarquização no OSM, a das rodovias em si. Neste caso, o tipo de raciocínio demandado na categorização mental dessas feições é mais concreto e

característico do nível de abstração inferior, cujo nível de categorização dentro da taxonomia é mais específico. Sob este aspecto, é importante pensar em quais atributos adicionais seriam acrescentados às rodovias para se fazer a diferenciação entre as rodovias primárias, secundárias e terciárias no processo de categorização mental. Isto quer dizer que esta categorização tem também baixa validade de sugestão porque compartilham muitos atributos com as rodovias ou estradas (sem entrar no mérito da diferenciação dos termos) em geral (ROSCH, 1973; ROSCH *et al.*, 1976; TVERSKY & HEMENWAY, 1984; MACEACHREN, 1995).

Sob a luz desses conceitos e análises, é possível afirmar que o alinhamento semântico foi mais efetivo para os termos de feições cujo nível de abstração é básico, que demandam um grau de raciocínio intermediário entre o concreto e o abstrato. Ou seja, termos de feições que possibilitam a diferenciação das feições mais próximas ao que se vê no mundo real: é trecho de via, que é diferente de edificação ou de área verde. Esse trecho de via pode ser diferenciado, mesmo com a maioria dos atributos comuns, em uma rodovia e um logradouro. O alinhamento semântico nas categorizações que foram além desse nível de abstração, seja mais abstrato ou mais concreto (não obstante outros fatores como a ambigüidade nas descrições semânticas), já demandam análises adicionais de contexto local e de uso das feições alinhadas.

Na tabela 6, também é possível observar que os dados classificados incorretamente tiveram porcentagens muito pequenas. Nesse caso, a possibilidade de a classificação original dos dados pelos colaboradores ter sido errada é maior. Para os trechos sem correspondência na base oficial, podem ser vias internas de condomínios, que não fazem parte do mapeamento oficial, novos trechos de vias ou trechos ainda não oficializados no banco de dados do IPPUC.

A tabela 7 mostra os resultados para segundo teste, realizado com os dados do município de Campo dos Goytacazes - RJ.

TABELA 7 – TESTE DE AVALIAÇÃO DO ALINHAMENTO CAMPO DOS GOYTACAZES

**TESTE DE AVALIAÇÃO DE QUALIDADE DO ALINHAMENTO - CAMPO DOS GOYTACAZES - RJ**

ETIQUETA (TAG)	Nº SEGMENTOS	CLASSIFICADO CORRETAMENTE	%	CLASSE ET-EDGV 3.0 ALINHADA	CLASSIFICADO INCORRETAMENTE trecho_arruamento trecho_rodoviário	%	CLASSIFICADO INCORRETAMENTE OUTROS		SEM CORRESPONDÊNCIA NA BASE OFICIAL	%
Residential	10722	8170	<b>76,20%</b>	trecho_arruamento	1019	9,50%	-	0%	1.533	14,30%
Unclassified	2560	1077	42,07%	trecho_arruamento	1130	<b>44,14%</b>	-	0%	353	13,79%
Trunk	462	454	<b>98,27%</b>	trecho_rodoviário	8	1,73%	-	0%		0,00%
Primary	721	708	<b>98,20%</b>	trecho_rodoviário	12	1,66%	-	0%	1	0,14%
Secondary	1151	621	53,95%	trecho_rodoviário	509	<b>44,22%</b>	-	0%	21	1,82%
Tertiary	1355	620	45,76%	trecho_rodoviário	729	<b>53,80%</b>	0	0%	6	0,44%
Motorway	0	0	0%	trecho_rodoviário	0	0%	0	0%	0	0%

FONTE: A autora (2020).

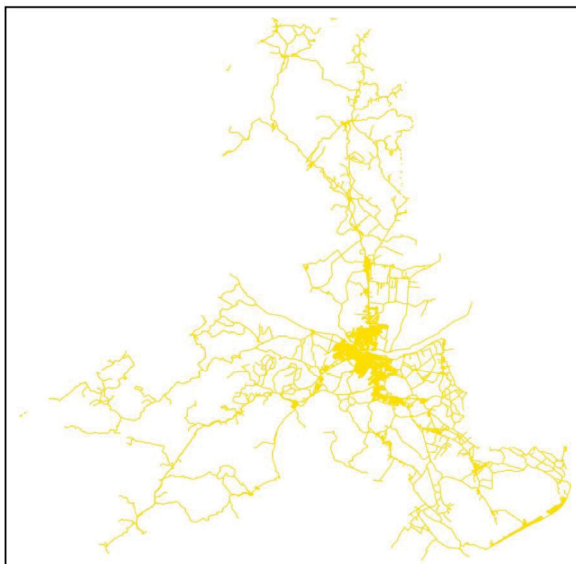
Na tabela 7 é possível observar que a melhor porcentagem dos segmentos classificados corretamente pelo alinhamento foi para as etiquetas *highway=trunk* (98%) e *highway=primary* (98%), que correspondem aos trechos rodoviários e rodovias e, para a etiqueta *highway=residential* (76%), que corresponde aos trechos de arruamento de vias residenciais. Nesse caso, a melhoria nas porcentagens da classificação das duas primeiras etiquetas, se comparadas às porcentagens do teste realizado para a cidade de Curitiba, se deve ao fato de a área do município de Campo dos Goytacazes ser maior e ter áreas rurais e urbanas. Para as demais etiquetas *highway=unclassified* (42%), *highway=secondary* (54%) e *highway=tertiary* (46%), as conclusões são as mesmas do teste para Curitiba. Vários fatores podem ter levado a essa baixa porcentagem dos segmentos classificados corretamente para essas etiquetas, as descrições semânticas ambíguas, necessidade de análise de contexto local e possíveis erros na classificação original dos dados pelos colaboradores. Em face dos conceitos de organização do conhecimento, as análises são as mesmas feitas para o teste de Curitiba-PR.

Na sequência são apresentadas três figuras com mapas esquemáticos apresenta das camadas de feições dos mapeamentos no teste de integração do município de Campo dos Goytacazes - RJ nas escalas 1:450.000 e 1:50.000 na tela.

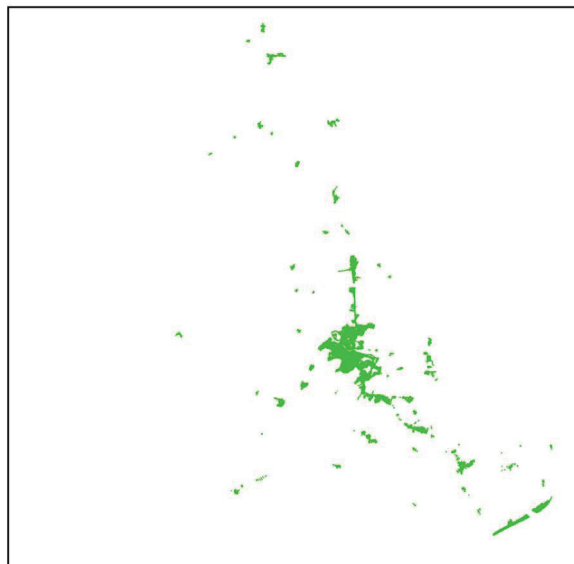
A figura 43 mostra as camadas de feições na escala 1:450.000.

FIGURA 43 – MAPA ESQUEMÁTICO DAS CAMADAS DE FEIÇÕES NA INTEGRAÇÃO ENTRE OS MAPEAMENTOS

**OSM 1:450.000**



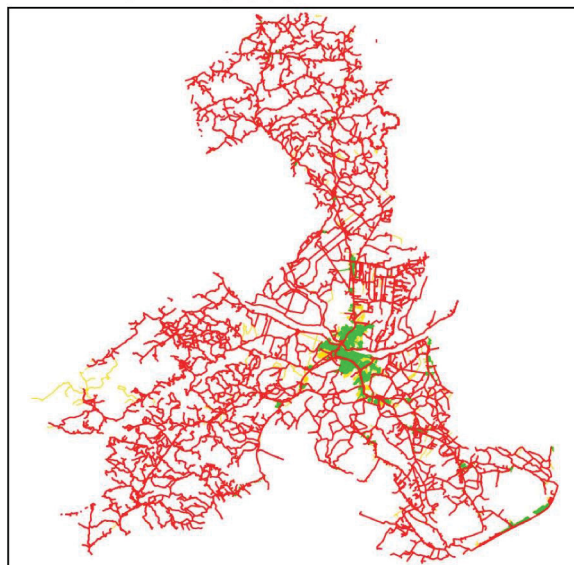
**ARRUAMENTO 1:450.000**



**RODOVIAS 1:450.000**



**HÍBRIDO 1:450.000**



FONTE: A autora (2020).

Nos mapas da figura 43 é possível observar que na base do município de Campo dos Goytacazes – RJ a separação entre arruamento e rodovia no mapeamento topográfico oficial produzido pelo IBGE é bem evidente. Os trechos de arruamento estão concentrados nas áreas urbanas, principalmente na parte central do município.

As estradas vicinais no mapeamento colaborativo do *OpenStreetMap* foram classificadas pelos colaboradores como vias residenciais (*highway=residential*) e na



base do mapeamento topográfico oficial foram classificadas como estradas. Isso ocorre porque embora essas vias sejam estradas vicinais, elas fornecem acesso às moradias rurais.

As figuras 44 e 45 mostram as camadas de feições na escala 1:50.000 na tela do computador.

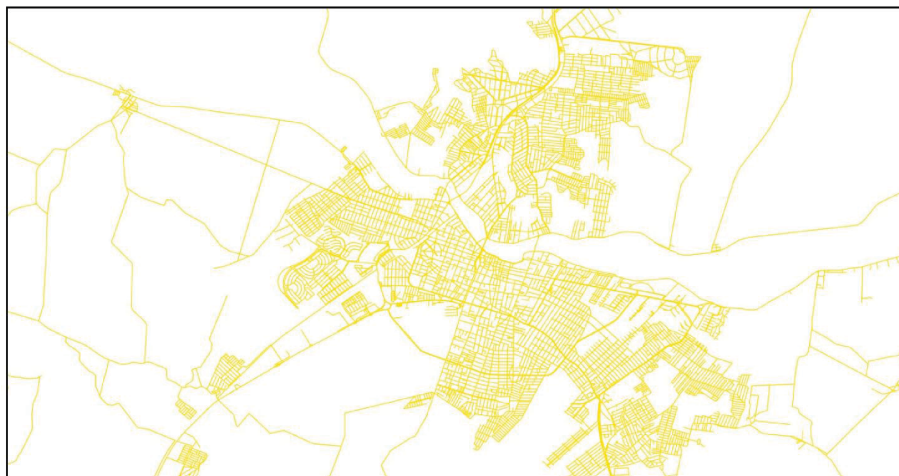
Nos mapas destas figuras é possível perceber que no mapeamento colaborativo do *OpenStreetMap* para o município de Campo dos Goytacazes – RJ, ocorre maior quantidade de feições em algumas áreas, possibilitando o uso destes dados em processos de atualização e completude da base do mapeamento topográfico oficial.

Também é evidente nos mapas esquemáticos das figuras 44 e 45, que a ocorrência dessas novas feições é maior nas áreas urbanas e menor nas áreas rurais, quando comparado com o mapeamento topográfico oficial produzido pelo IBGE. No caso deste município, tal fato confirma o observado na literatura de que a disponibilidade de dados do *OSM* é maior nas áreas urbanas (LUDWIG & ZIPF, 2019; ESTIMA & PAINHO, 2013) e que as colaborações são menores nas áreas rurais (CAMBOIM *et al.*, 2015).

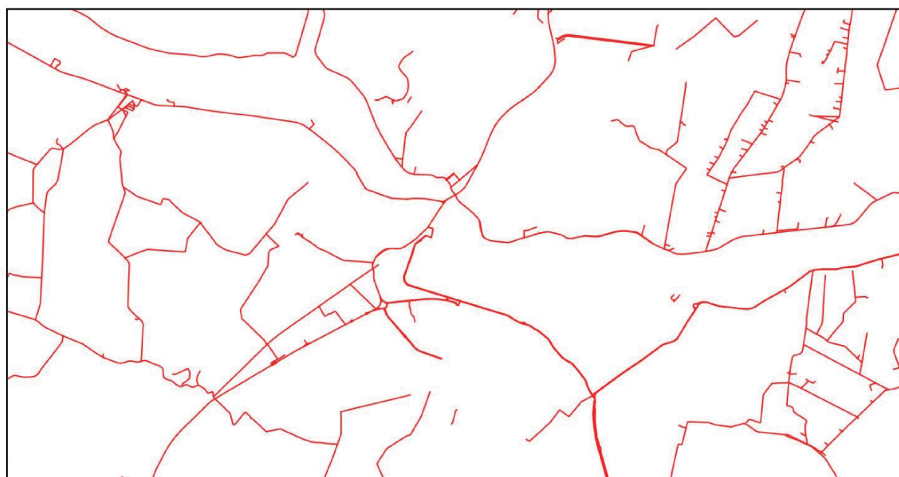


FIGURA 44 – MAPA ESQUEMÁTICO DAS CAMADAS DAS FEIÇÕES NA INTEGRAÇÃO ENTRE OS MAPEAMENTOS

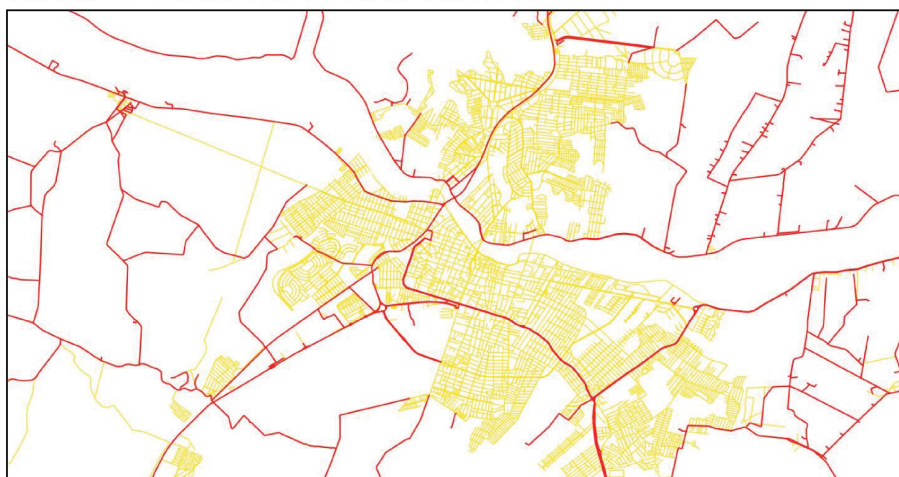
**OSM 1:50.000**



**RODOVIAS 1:50.000**



**HÍBRIDO OSM / RODOVIAS 1:50.000**



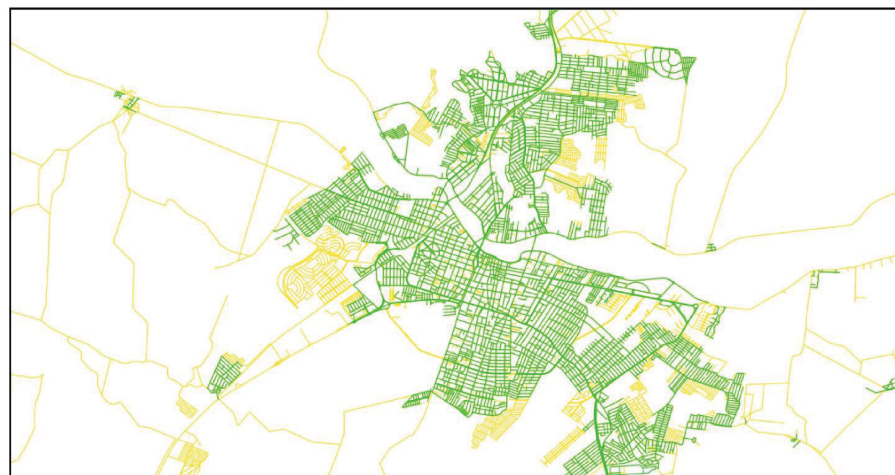
FONTE: A autora (2020).

FIGURA 45 – MAPA ESQUEMÁTICO DAS CAMADAS DAS FEIÇÕES NA INTEGRAÇÃO ENTRE OS MAPEAMENTOS

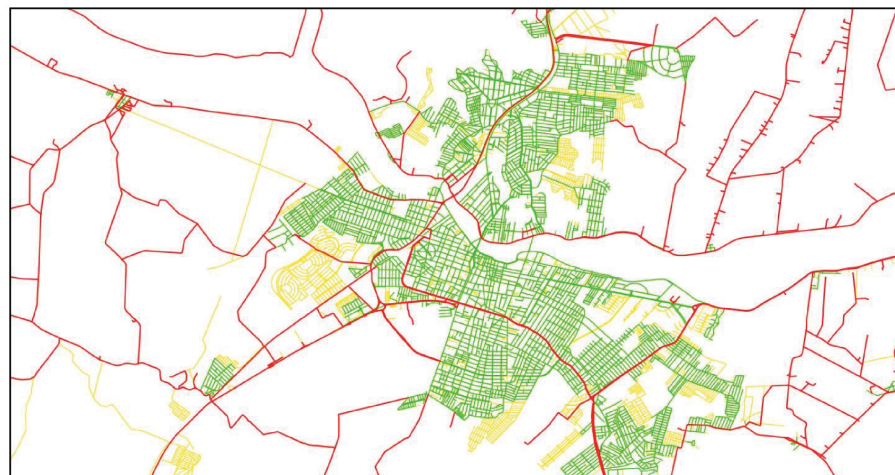
**ARRUAMENTO 1:50.000**



**HÍBRIDO OSM / ARRUAMENTO 1:50.000**



**HÍBRIDO 1:50.000**



FONTE: A autora (2020).

O segundo mecanismo de avaliação da ontologia foi a classificação quanto ao nível de dificuldade do alinhamento semântico. As tabelas 8, 9 e 10 apresentam os resultados desta classificação.

Analisando-se os resultados, de modo geral o nível de dificuldade do alinhamento semântico foi considerado baixo e médio. Analisando-se individualmente algumas categorias da ET-EDGV 3.0 com as respectivas chaves do *OpenStreetMap*, como no caso da categoria Energia e Comunicações, observou-se que embora as feições tenham descrições semânticas com conteúdo bastante técnico, a dificuldade de classificação foi média, pois muitas dessas descrições possuíam palavras-chave que permitiram o alinhamento. Nos casos em que houve dúvidas, essas foram sanadas com consultas à Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais (ET-ADGV 3.0).

Para a categoria Cultura e Lazer, ocorreram casos nos quais as etiquetas do *OSM*, por exemplo, utilizavam o tipo de esporte para descrever o tipo de pista no qual o esporte é praticado, ou seja, descrições ambíguas que levaram a múltiplas correspondências. Outra situação, que ocorreu não apenas para essa categoria, mas também para muitas outras é a ocorrência de etiquetas diferentes, às vezes até três, com o mesmo significado.

As categorias que tiveram maior nível de dificuldade foram Sistema de Transporte Rodoviário e Edificações, cujo motivo principal é a ambigüidade nas descrições semânticas. Em algumas classes como no caso de Edif\_Saude a descrição semântica na ET-EDGV 3.0 é genérica com a distinção através do valor de domínio do atributo ClasseAtivEcon para distinguir a atividade. Por exemplo, atributo ClasseAtivEcon com valor de domínio atendimento hospitalar (hospital). No caso do *OSM* as diferenciações são diretas como *amenity=clinic*, *amenity=doctors*, *amenity=hospital*, *amenity=dentist*.

Na ET-EDGV 3.0 ocorrem algumas distinções minuciosas, pelo fato de se tratar de mapeamento oficial, como no caso de residencial e habitacional, desportivo e recreativo. No modelo conceitual do *OpenStreetMap*, as descrições utilizam essas palavras como sinônimos, dificultando o alinhamento devido a ambigüidade.

TABELA 8 – CLASSIFICAÇÃO EM RELAÇÃO AO NÍVEL DE DIFICULDADE DO ALINHAMENTO SEMÂNTICO

## CLASSIFICAÇÃO EM RELAÇÃO AO NÍVEL DE DIFICULDADE DO ALINHAMENTO SEMÂNTICO

MAPEAMENTO TOPOGRÁFICO EM PEQUENAS ESCALAS		
DIFICULDADE BAIXA	DIFICULDADE INTERMEDIÁRIA	DIFICULDADE ALTA
	<b>Energia e Comunicações</b>	
	power	
building: condition		
	man_made	
	man_made= tower tower_type= *	
	building	
	<b>Estrutura Econômica</b>	
industrial		
building:material		
landuse		
	<b>Hidrografia</b>	
	natural	
	natural=water	
	water=*	
	<b>Limites e Localidades</b>	
leisure		
boundary		
landuse		
building: condition		
	<b>Relevo</b>	
natural		
	<b>Saneamento Básico</b>	
	natural= water	
	water=*	
		amenity
	<b>Sistema de Transporte</b>	
	public_transport	
amenity		
	building	
building:condition		
bridge		
man_made		
tunnel:		
	surface	
railway		
name		
		operator:type
aeroway		
bridge:structure		
		highway
	<b>Sistema de Transporte / Aeroportuário</b>	
aeroway		
building:condition		
	<b>Sistema de Transporte / Dutos</b>	
tunnel		
man_made		
building:condition		
landuse		

FONTE: A autora (2020).

TABELA 9 – CLASSIFICAÇÃO EM RELAÇÃO AO NÍVEL DE DIFICULDADE DO ALINHAMENTO SEMÂNTICO (cont.)

**CLASSIFICAÇÃO EM RELAÇÃO AO NÍVEL DE DIFICULDADE DO ALINHAMENTO SEMÂNTICO**

<b>MAPEAMENTO TOPOGRÁFICO EM PEQUENAS ESCALAS</b>		
<b>DIFICULDADE BAIXA</b>	<b>DIFICULDADE INTERMEDIÁRIA</b>	<b>DIFICULDADE ALTA</b>
	<b>Sistema de Transporte / Ferroviário</b>	
landuse		
	railway	
		public_transport
		operator:type
building:condition		
	gauge	
electrified		
tracks		
	<b>Sistema de Transporte / Hidroviário</b>	
		landuse
natural= water		
water=*		
	<b>Sistema de Transporte / Rodoviário</b>	
		highway
		operator:type
		surface
building:condition		
		lanes
		key:ref
maxspeed		
	<b>Vegetação</b>	
		landuse
		natural
<b>MAPEAMENTO TOPOGRÁFICO EM GRANDES ESCALAS</b>		
<b>DIFICULDADE BAIXA</b>	<b>DIFICULDADE INTERMEDIÁRIA</b>	<b>DIFICULDADE ALTA</b>
	<b>Área Verde</b>	
	landuse	
natural		
<b>Classes Base do Mapeamento Topográfico em Grandes Escalas</b>		
	landuse	
name		
animal		
	amenity	
	religion	
	railway	
	power	
	building	
	industrial	
	man_made	
bridge		
	bridge:structure	
		highway
		lanes
		key:ref
maxspeed		
tunnel		

FONTE: A autora (2020).



TABELA 10 – CLASSIFICAÇÃO EM RELAÇÃO AO NÍVEL DE DIFICULDADE DO ALINHAMENTO SEMÂNTICO (cont.)

CLASSIFICAÇÃO EM RELAÇÃO AO NÍVEL DE DIFICULDADE DO ALINHAMENTO SEMÂNTICO		
MAPEAMENTO TOPOGRÁFICO EM GRANDES ESCALAS		
DIFICULDADE BAIXA	DIFICULDADE INTERMEDIÁRIA	DIFICULDADE ALTA
	<b>Cultura e Lazer</b>	
leisure		
sport		
building:condition		
name		
highway		
		operator:type
	landuse	
tourism		
	amenity	
	<b>Edificações</b>	
building		
name		
building:condition		
	building:material	
high		
tourism		
		operator:type
		addr:
leisure		
amenity		
	office	
		industrial
		public_transport
	aeroway	
	natural= water	
	water=*	
	amenity=recycling	
	recycling_type=*	power
landuse		
	<b>Estrutura de Mobilidade Urbana</b>	
		highway
building:condition		
	surface	surface
	aerialway	
	ramp	
		public_transport

FONTE: A autora (2020).

O terceiro mecanismo de avaliação da qualidade da ontologia foi a aplicação do *checklist* para detecção de inconsistências na ontologia. O resultado dessa avaliação encontra-se na tabela 11. Nessa tabela são listadas as possíveis inconsistências, e se estas foram detectadas.

TABELA 11 – DETECÇÃO DE INCONSISTÊNCIAS NA COMPATIBILIZAÇÃO

**DETECÇÃO DE INCONSISTÊNCIAS**

1. Conceitos com um mesmo significado, mas rotulados com nomes diferentes.	Sim
2. Conceitos rotulados com o mesmo nome, mas com significados diferentes.	Não
3. Diferenças na escrita dos termos das ontologias como, por exemplo: um termo no plural e outro no singular, um no feminino e outro no masculino, e em diferentes tempos verbais.	Sim
4. Propriedades com um mesmo significado, mas rotuladas com diferentes nomes.	Não
5. Propriedades rotuladas com o mesmo nome, mas com significados diferentes.	Não
6. Diferenças nas restrições e nas propriedades utilizadas.	Sim
7. Diferenças nas propriedades utilizadas em restrições - conceitos relacionados que são similares.	Não
8. Diferenças nos conceitos relacionados utilizados em restrições - propriedades que são similares.	Não
9. Diferenças no número de restrições (diferenças nos casos onde exista uma interseção de restrições).	Não
10. Verificação se todos os conceitos a que as propriedades se relacionam são equivalentes em ambas as ontologias comparadas.	Não
11. Verificação se todos os conceitos que utilizam uma propriedade são consistentes em ambas as ontologias comparadas.	Não

FONTE: Adaptado de Felícicimo (2004).

Sobre a primeira inconsistência “conceitos com um mesmo significado, mas rotulados com nomes diferentes”, ocorreram diversos casos, mas esses não inviabilizaram o processo de alinhamento.

Esse aspecto já era previsto, pois a ontologia do mapeamento topográfico oficial nacional é de nível local e a do mapeamento colaborativo é de nível global. O modelo conceitual do *OSM* tem as traduções dos seus elementos em diversos idiomas, entretanto, as chaves e etiquetas permanecem em inglês, o idioma nativo da plataforma. A tradução da descrição semântica das chaves e etiquetas é feita juntamente com as descrições dos seus significados, então alguns termos possuem traduções com termos não usuais. Exemplos: caixa d’água na ET-EDGV 3.0 e *man\_made=water\_tower*, traduzido como castelo d’água, depósito de água elevado ou reservatório de água elevado no modelo conceitual do *OpenStreetMap*.

Outro caso que ocorre no *OSM*, é que existem etiquetas que têm o valor igual e as chaves diferentes. Esse é um dos aspectos que torna o modelo conceitual da plataforma colaborativa flexível e aberto. Entretanto, esse aspecto pode gerar mal entendido na interpretação ou redundâncias na classificação. Como exemplo cita-se *amenity=hotel* e *tourism=hotel*. Outro exemplo, é o valor de domínio salina do atributo tipoExtMin da classe ExtMineral e as etiquetas *landuse=salt\_pond* e *industrial=salt\_pond*. A descrição semântica de *industrial=salt\_pond* é: “Salina é uma área de produção de sal marinho pela evaporação da água do mar ou de lagos de



água salgada. É uma alternativa a etiqueta *landuse=salt\_pond* mais amigável para os consumidores de dados”. A descrição semântica de *landuse=salt\_pond* é: “Salina é um local para produção de sal marinho através da inundação de ‘campos’ com água salgada e posterior evaporação da água através de exposição solar”.

Sobre a terceira inconsistência, “diferenças na escrita dos termos das ontologias como, por exemplo: um termo no plural e outro no singular, um no feminino e outro no masculino, e em diferentes tempos verbais”, ocorreram diversos casos, mas não interferiram no processo de alinhamento. Exemplo: o atributo revestimento com valor de domínio “pavimentado” e a etiqueta *surface=paved* superfície “pavimentada” no *OSM*, ou seja, um termo no masculino e outro no feminino.

Sobre a sexta inconsistência, “diferenças nas restrições e nas propriedades utilizadas”, devido ao fato de os modelos conceituais serem diferentes, propriedades, como por exemplo, as geométricas têm diferenças conceituais. No caso da ET-EDGV 3.0 uma feição pode ser representada pelas três geometrias previstas nos casos de generalização ou, caso as instâncias sejam representadas pela agregação de instâncias de classes de objetos com diferentes primitivas geométricas. No modelo do *OSM*, podem ocorrer diferentes geometrias para a mesma feição, mas não existe o conceito de generalização e agregação. No *OSM* as diferentes geometrias para uma mesma feição podem ser resultado de contribuições em diferentes níveis de escala. Por exemplo, feições importadas de um banco de dados ou feições desenhadas sobre uma imagem de satélite, nas quais uma pode ser mais detalhada resultando em uma geometria de caminho fechado (área) e outra em um nó (ponto).

As demais inconsistências do *check list* não foram detectadas. De modo geral, embora tenham sido encontradas algumas inconsistências, essas não inviabilizaram a estratégia de alinhamento entre as ontologias.

## 5. CONCLUSÕES

Nesta pesquisa foi proposto um método de compatibilização semântica entre o modelo conceitual do mapeamento colaborativo da plataforma *OpenStreetMap* (*OSM*) e a Especificação Técnica para Estruturação de Dados Geoespaciais

Vetoriais (ET-EDGV 3.0), o modelo conceitual do mapeamento topográfico oficial brasileiro através de uma ontologia que alinha os dois modelos.

Sob o ponto de vista científico a pesquisa vai ao encontro das pesquisas nacionais e internacionais (Cheatham *et al.*, 2019; Ludwig & Zipf, 2019; Hagedorn *et al.*, 2019; Novack *et al.*, 2018; Castañeda Filho, 2017; Estima & Painho, 2013; Al-Bakri & Fairbairn, 2012; Baglatzi, *et al.*, 2012; PINHO & GOLTZ, 2003; Noy & Musen, 2003; Lima *et al.*, 2002; Noy & Musen, 2001) com foco na compatibilização semântica, na sua implementação através de ontologias e na integração entre o mapeamento colaborativo e o oficial.

Toda a metodologia desenvolvida foi fundamentada na literatura existente e sua documentação em linguagem natural e gráfica gerou uma base de conhecimento que consiste em uma das principais contribuições desta pesquisa. Essa base de conhecimento permite acessar o conhecimento implícito nas duas ontologias e conectá-las de modo a permitir a interoperabilidade semântica e lógica entre os bancos de dados dos dois mapeamentos

Na literatura os alinhamentos têm sido feitos apenas no primeiro nível taxonômico (LUDWIG & ZIPF, 2019) ou, no primeiro e no segundo nível taxonômico da estrutura conceitual dos mapeamentos (ESTIMA & PAINHO, 2013). O alinhamento semântico proposto nesta pesquisa traz como um avanço o fato de ter sido realizado em todos os níveis taxonômicos dos modelos conceituais, do mais alto ao mais baixo. Este aspecto permitiu obter uma visão geral e completa das estruturas conceituais e apontar categorias de dados com maior e menor potencial de integração entre os mapeamentos.

Como avanços para a área de Geoinformação a ontologia concebida nesta pesquisa contribui para o desenvolvimento de sistemas de integração entre mapeamentos, para a geração de novas ferramentas de inteligência artificial e sistemas baseados em conhecimento, além da geração de recursos para exploração e busca dos dados referentes aos mapeamentos na *internet*. Estas são ferramentas, que se aplicadas, trarão impactos sociais, culturais e econômicos às cidades e ao país como um todo.

Entende-se que os objetivos da pesquisa foram alcançados. Entre os resultados, o desenvolvimento do *plugin* que utiliza a base de conhecimento gerada pelo alinhamento semântico demonstrou a aplicabilidade da solução. Os resultados dos testes efetuados com o *plugin* para as duas áreas de estudo no alinhamento das

classes trecho rodoviário e trecho de arruamento da categoria de Sistema de Transporte Rodoviário da ET-EDGV 3.0 com as etiquetas da chave *highway* do OSM, demonstraram que o alinhamento produziu boas correspondências. Concluiu-se que existe influência da organização mental do conhecimento nas classificações e consequentemente no alinhamento semântico. O alinhamento semântico gerado foi eficaz na classificação de dados nos níveis de abstração superior e básico. Nos casos cujo nível de detalhamento e abstração é maior, ou seja, nos níveis taxonômicos inferiores ou subordinados, a análise de contexto local e propósito de uso dos dados tornam-se necessários à classificação final dos dados. Ao mesmo tempo, este aspecto remete a uma limitação do método desenvolvido, pois não permite a automatização total na integração dos bancos de dados sem que haja a análise de contexto local e de uso das informações. Nesse sentido, alinhamentos desse tipo podem ajudar no desenvolvimento de novas ferramentas para a classificação de dados entre diferentes fontes.

Concluiu-se ainda que a integração dos dados do *OpenStreetMap* aos dados oficiais se mostrou promissora e que algumas categorias de informações podem ser mais facilmente integradas do que outras, devido a questões legais e da organização mental do conhecimento na categorização das feições. De acordo com os resultados obtidos no alinhamento semântico, as categorias que demonstraram maior potencial para serem utilizadas na integração entre os mapeamentos foram as dos Sistemas de Transporte (Aeroportuário, Dutos, Ferroviário e Rodoviário), Energia e Comunicações, Estrutura Econômica, Estrutura de Mobilidade Urbana e Edificações. As categorias cujo potencial de integração demonstrou não serão promissor foram Limites e Localidades, Vegetação, Relevo, Hidrografia, Sistema de Transporte/Hidroviário, Saneamento Básico, Cultura e Lazer e, Classes Base do Mapeamento Topográfico em Grandes Escalas.

Como lições aprendidas sobre os modelos conceituais dos mapeamentos, destaca-se que o modelo conceitual do *OpenStreetMap* é muito bem documentado, ilustrado e dinâmico. A utilização da *Extensive Markup Language* (XML) para estruturar o modelo conceitual confirma as vantagens da sua utilização e a crença de que esta linguagem tem potencial para ser uma linguagem universal para representação de dados conforme relatado na literatura (SOUSA *et al.*, 2018; MORAIS & AMBRÓSIO, 2007; LIMA *et al.*, 2002).

A estrutura do OSM é uma prova contemporânea de aplicação da inteligência coletiva, dado o modo como é feita a gestão das etiquetas e as suas traduções em todo o mundo pela comunidade de contribuidores, bem como, a gestão de toda a estrutura conceitual, organizacional e do mapeamento como um todo dessa plataforma colaborativa.

Como padrão de modelo conceitual para bancos de dados geoespaciais com o objetivo de armazenar e manipular as bases cartográficas, a ET-EDGV 3.0 preenche uma lacuna importante no cenário da geoinformação no país. A especificação cumpre seu papel na viabilização do compartilhamento de dados, na interoperabilidade e futuramente na racionalização de recursos entre os produtores e usuários de dados e informação cartográfica para as atividades de planejamento no Brasil.

Entretanto, questiona-se a sua representação através do modelo OMT-G. Primeiramente, o modelo é extenso e rígido, com muitos elementos obrigatórios, os quais o mapeamento brasileiro não tem insumos para preencher. A OMT-G não é uma linguagem universal, na contramão da tendência de utilizar modelos globais como é o caso a especificação ISO 19125:2004 *Simple Feature Access* (ISO, 2004) que vem sendo utilizada em iniciativas internacionais, a exemplo da *Infrastructure for Spatial Information in Europe* (INSPIRE) (MASÓ *et al.*, 2019; STEPHEN & HAHMANN, 2019; ERIKSSON *et al.*, 2018).

Entre os problemas encontrados, as definições de algumas classes, por vezes dão margem para interpretações equivocadas ou não englobam alguns elementos importantes, como é o caso dos paraciclos que poderiam fazer parte da classe Estacionamento ou ter uma classe específica. Outro exemplo é a definição da classe Quadra: “uma composição de lotes se caracteriza como uma unidade básica componente de um bairro” (DSG, 2017); contudo a especificação não tem a classe Lote. Os lotes fazem parte do mapeamento cadastral, e provavelmente a definição de quadra deveria ser corrigida nas próximas versões da ET-EDGV. Outra questão é que alguns valores de domínio assumidos pelos atributos não têm as suas descrições semânticas, situação que no caso do alinhamento com as etiquetas do OSM não permitiu a compatibilização, dado que muitas correspondências se deram nesse nível. Também existem alguns erros textuais nos valores de domínio associados aos atributos que podem levar a interpretações equivocadas, como no caso da classe Ciclovía. Cabe salientar, que essas questões podem ser resolvidas,

deixando a ET-EDGV em constante processo de atualização e discussão, pois se trata de uma norma que deve ser aperfeiçoada continuamente devido à sua importância para o mapeamento nacional. Para tanto, é importante manter e fortalecer os órgãos que congregam usuários e produtores de mapeamento no Brasil como é o caso da Comissão Nacional de Cartografia (CONCAR).

Por fim, intenciona-se que as soluções geradas nesta pesquisa sejam compartilhadas e aprimoradas colaborativamente, sendo um ponto de partida para a discussão sobre a integração semântica entre modelos conceituais como uma possível solução para a interoperabilidade de dados. O universo de pessoas envolvidas nessa discussão abrange os produtores de mapeamento oficial em diversas escalas, que incluem as entidades dos governos estaduais e federal, as 5.560 prefeituras brasileiras, empresas, e outros, além dos mais de cinco milhões de colaboradores globais do *OpenStreetMap*, ou seja, grandes comunidades geradoras de dados, que pela natureza de uso geral dos mapas topográficos atinge uma grande e diversa população. Sendo assim, a questão de aprimoramento colaborativo tem um grande potencial inclusive no meio acadêmico.

## 6. PESQUISAS FUTURAS

Para pesquisas futuras, sugere-se a realização de investigações mais aprofundadas a fim de obter soluções para as discrepâncias conceituais e semânticas entre as duas ontologias, principalmente em relação aos casos de descrições ambíguas que conduzem a múltiplas correspondências entre as classificações. Ainda sob esse aspecto, sugere-se o desenvolvimento de metodologias para avaliar a acurácia dos alinhamentos.

Com a base de conhecimento gerada pelo alinhamento sugere-se novas pesquisas para o desenvolvimento de ferramentas de aprendizagem de máquina para derivar soluções semânticas entre as descrições. Outras pesquisas podem ainda, desenvolver ferramentas de uso de ontologias para facilitar o compartilhamento de dados entre os modelos.

Recomenda-se ainda, estudos sobre as questões relativas à escala nos modelos conceituais, sobre a generalização semântica dos conceitos, sobre os regionalismos lingüísticos e em linguagem múltiplas, com traduções para outros

idiomas como o espanhol, por exemplo, para suplantar bases de dados sulamericanas. Ainda sob o aspecto do idioma, sugere-se pesquisas sobre o impacto do idioma nas descrições semânticas (*crosslinguistic influence*).

Quanto à organização mental do conhecimento na categorização das feições, considera-se interessante investigar se existe diferença nas classificações das feições em diferentes regiões e se elas variam no tempo.

Indica-se adicionalmente, a formalização da ontologia através de uma linguagem computacional e estudos para geração de ferramentas de manutenção da ontologia. Nesse contexto, recomenda-se ainda a produção de material de treinamento e divulgação da ontologia para expandir a rede de usuários envolvidos nestas discussões.

Por fim, sugere-se pesquisas complementares sobre as questões legais e éticas a respeito integração entre os dados oficiais e colaborativos.

## 7. REFERÊNCIAS

AGARWAL, P. Ontological considerations in GIScience. **International Journal of Geographical Information Science**, n.19, p. 501-536, 2005.

AHN, L. VON & DABBISH, L. Labeling images with a computer game. **In: Proceedings of the SIGCHI conference on Human factors in computing systems**, ACM, 2004. p. 319-326.

AL-BAKRI, M. & FAIRBAIRN, D. Assessing similarity matching for possible integration of feature classifications of geospatial data from official and informal sources. **International Journal of Geographical Information Science**, 26:8, p.1437-1456, 2012. Disponível em: <http://dx.doi.org/10.1080/13658816.2011.636012>. Acesso em: 05/05/2020.

ALMEIDA, M & BAX, M. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Revista Ciência da Informação**, v.32, n.3, 2003.

ARENHART, J. R. B. & KRAUSE, D. Semântica Quase-conjuntista e Compromisso Ontológico. **Cognitio**, São Paulo, v. 10, n. 2, p. 191-207, jul./dez. 2009.

ARIZA-LÓPEZ, F. J. A. **Calidad en la producción cartográfica**. Ra-Ma, Madrid, 389 p., 2002.

ARIZA-LÓPEZ, F. J.; RODRÍGUEZ AVI, J. Aplicación de las normas ISO 2859-1 e ISO 2859-2 en el control posicional de suministros de datos espaciales, aplicación al caso de puntos y líneas. **GeoFocus**, v.15, p. 27-46, 2015.

BAGLATZI, A.; KOKLA, M.; KAVOURAS, M. Semantifying OpenStreetMap. **The 11th International Semantic Web Conference**, 2012. p. 39-48.

BALLATORE, A. BERTOLOTTO, M., WILSON, D.C. The semantic similarity ensemble. **Journal of Spatial Information Science**, n.7, p. 27-44, 2014.



BEARDEN, M. J. The National Map Corps: The USGS' Volunteer Geographic Information Program. **VGI Workshop (NCGIA)**, Santa Barbara, California, 2007. Disponível em: <[www.ncgia.ucsb.edu/projects/vgi/docs/position/Bearden\\_paper.pdf](http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Bearden_paper.pdf)> Acesso em: 30/10/2017.

BENNETT, J. **OpenStreetMap: Be Your Own Cartographer**. Packt Publishing, Birmingham, UK, 2010, 234 páginas.

BERLIN, B. Speculations on the Growth of Ethnobotanical Nomenclature. **Language in Society**, v.1, n.1, p. 51-86, April 1972. Disponível em: <https://doi.org/10.1017/S0047404500006540>. Acesso em: 06/05/2020.

BERLIN, B.; BREEDLOVE, D. E. ; PETER H. RAVEN, P.H. General Principles of Classification and Nomenclature in Folk Biology. **American anthropologist**, v. 75, n.1, p. 214-242, 1973.

BERNERS-LEE, T., HENDLER, J. e LASSILA, O. The semantic web. **Scientific American**, n. 284, p. 28-37, 2001.

BOARD, C. Map reading tasks appropriate in experimental studies in cartographic communication. **The Canadian Cartographer**, v. 15, 1978.

BORGES, K. A. V.; DAVIS JR., C. A.; LAENDER, A. H. F. Modelagem conceitual de dados geográficos. In: CASANOVA, M. A.; CÂMARA, G.; DAVIS JR., C. A.; Lúbia VINHAS, L.; QUEIROZ, G. (ed). **Bancos de Dados Geográficos**. Curitiba, Editora MundoGEO, 2005. Disponível em: <<http://www.dpi.inpe.br/livros/bdados/>>. Acesso em: 30/10/2017.

BORST, W. Construction of Engineering Ontologies for Knowledge Sharing and Reuse. PhD thesis, University of Twente, P.O. Box 217 - 7500 AE Enschede - The Netherlands, 1997.

BORTOLINI, E.; SILVA, L.S.L.; MACHADO, A. A.; PAIVA, C. A.; CAMBOIM, S. P. Potenciais Categorias de Informações Geográficas do Mapeamento Colaborativo para o Mapeamento Oficial. **Anais do X Colóquio Brasileiro de Ciências Geodésicas**. Curitiba, 06-08 de junho de 2018.

BOUQUET, P.; SERAFINI, L.; ZANOBINI, S. Semantic Coordination: A New Approach and an Application. In: Fensel D., Sycara K., Mylopoulos J. (eds) **The Semantic Web – ISWC**, 2003. Lecture Notes in Computer Science, v. 2870. Springer, Berlin, Heidelberg, p. 130-143.

BRACHMAN, R. J. & LEVESQUE, H. J. What makes a knowledge base knowledgeable? A view of databases from the knowledge level. **Proceedings from the first international workshop on Expert database systems**. January, 1986. p. 69-78.

BRASIL. **Decreto 243, de 28 de Fevereiro de 1967**. Fixa as Diretrizes e Bases da Cartografia Brasileira. Disponível em: <http://www.inde.gov.br/legislacao.html>. Acesso em: 11/12/2017.

BRASIL. **Decreto 89.817, de 20 de junho de 1984**. Estabelece as Instruções Reguladoras das Normas Técnicas da Cartografia Nacional. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/decreto/1980-1989/D89817.htm](http://www.planalto.gov.br/ccivil_03/decreto/1980-1989/D89817.htm). Acesso em: 11/12/2017.

BRASIL. **[Constituição (1988)]. Constituição da República Federativa do Brasil**. In: Vade Mecum Acadêmico de Direito, Editora Rideel, 8ª edição. São Paulo, 2009.

BRASIL. **Decreto nº 6.666, de 28 de Novembro de 2008**. Institui, no âmbito do Poder Executivo federal, a Infraestrutura Nacional de Dados Espaciais (INDE), e dá outras providências. Brasília, Diário Oficial da União, 2008. Disponível em: <http://www.inde.gov.br/legislacao.html>. Acesso em: 25/11/2019.

BRASIL. **Lei nº 12.527, de 18 de novembro de 2011**. Regula o acesso a informações previsto no inciso XXXIII do art. 5º da Constituição Federal e dá outras

providências. Disponível em: [www.planalto.gov.br/ccivil\\_03/\\_ato20112014/2011/lei/l12527.htm](http://www.planalto.gov.br/ccivil_03/_ato20112014/2011/lei/l12527.htm). Acesso em: 28/12/2017.

BRASIL. **Lei nº 9.503 de 23 de setembro de 1997**. Institui o Código de Trânsito Brasileiro. 7ª Ed.

BRASIL. **Lei Complementar nº 116, de 31 de julho de 2003**. Dispõe sobre o Imposto Sobre Serviços de Qualquer Natureza, de competência dos Municípios e do Distrito Federal, e dá outras providências. Diário Oficial da União, Brasília, DF, Seção 1, p. 3.

BRASIL. **Lei nº 12.379 de 6 de janeiro de 2011. Dispõe sobre o Sistema Nacional de Viação – SNV**; ...; e dá outras providências. Diário Oficial da União, Brasília, DF.

BRASIL. CASA CIVIL – SUBCHEFIA PARA ASSUNTOS JURÍDICOS. **Decreto nº 6.666 de 27/11/2008, institui a Infraestrutura Nacional de Dados Espaciais – INDE**. Disponível em: [http://www.inde.gov.br/images/inde/20@Decreto6666\\_27112008.pdf](http://www.inde.gov.br/images/inde/20@Decreto6666_27112008.pdf). Acesso em: 04/11/2017.

BRASIL. **Lei nº 12.651, de 25 de maio de 2012. Código Florestal Brasileiro**. Dispõe sobre a proteção da vegetação nativa; (...) e dá outras providências. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato20112014/2012/lei/l12651.htm](http://www.planalto.gov.br/ccivil_03/_ato20112014/2012/lei/l12651.htm). Acesso em: 26/01/2020.

BRAVO, J. V. M. **A Confiabilidade Semântica das Informações Geográficas Voluntárias como Função da Organização Mental do Conhecimento Espacial**. Dissertação de Mestrado. 139 p. Universidade Federal do Paraná, Programa de Pós-Graduação em Ciências Geodésicas, Curitiba (PR), 2014.

BRAVO, V. M. B.; CAMBOIM, S. P.; MENDONÇA, A. L. A.; SLUTER, C. R. A compatibilidade dos metadados disponíveis em sistemas VGI com o perfil de metadados empregado na Infraestrutura Nacional de Dados Espaciais do Brasil (INDE-BR). **Boletim de Ciências Geodésicas**, versão *online*, Curitiba, v. 21, nº 3, p. 465-483, jul-set, 2015. ISSN 1982-2170.

BRIN, S. & PAGE, L. Reprint of: The anatomy of a large-scale hypertextual Web search engine. **Computer Networks**, v. 56, n. 18, p. 3825-3833, 2012.

BROWN, R. How shall a thing be called? **Psychological Review**, v.65, n.1, 1958.

CAMBOIM, S. P. & SLUTER, C. R. The National Topographic Mapping as an Indispensable Database for a Brazilian National Spatial Data Infrastructure (NSDI). **Proceedings of the 24th International Cartographic Conference**, Santiago de Chile, Chile, 15-21 November, 2009.

CAMBOIM, S. P.; BRAVO, J. V. M.; SLUTER, C. R. An investigation into the completeness of, and updates to, the Open Street Map data in a heterogeneous area in Brazil. **ISPRS International Journal of Geo-Information**, v. 4, n. 3, p. 1366-1388, 2015.

CARISSIMI, E.; ETGES, V. E.; COLLISCHONN, E. Diagnóstico da Cartografia no Brasil e no Rio Grande do Sul frente ao Paradigma Geotecnológico e o Desenvolvimento Regional. **Comissão Econômica para a América Latina e o Caribe (CEPAL) – 60 anos de Desenvolvimento na América Latina**, Santa Cruz do Sul, RS, Brasil, 17 a 19 de agosto de 2011. Disponível em: <<http://www.unisc.br/site/sidr/2011/textos/51.pdf>>. Acesso em: 16/12/2017.

CASTAÑEDA FILHO, R. M. **Uma Ontologia de Referência para o Domínio das Divisões e Limites Geográficos Territoriais do Brasil**. Tese de doutoramento. 633p. Universidade Federal do Rio de Janeiro, Programa de Pós-Graduação em Geografia, Rio de Janeiro, 2017.

CHEATHAM, M., VARANKA, D., ARAUZ, F., ZHOU, L. Alignment of surface water ontologies: a comparison of manual and automated approaches. **Journal of Geographical Systems**, v. 22, p. 267–289, 2019.

CHUJAI, P., KERDPRASOP, N., e KERDPRASOP, K. On transforming the ER model to ontology using Protégé OWL tool. **International Journal of Computer Theory and Engineering**, v.6, p. 484-489, 2008.

COAST, S. **The Book of OSM**. CreateSpace Independent Publishing Platform, 2015. 290 p. ISBN-13: 978-1514232743.

CODESCU, M.; HORSINKA, G.; KUTZ, O.; MOSSAKOWSKI, T.; RAU, R. OSMonto - An Ontology of OpenStreetMap Tags. **State of the Map Europe (SOTM-EU)**, 2011.

COLEMAN, D. J. Potential Contributions and Challenges of VGI for Conventional Topographic Base-Mapping Programs. In: SUI, D.; ELWOOD, S.; GOODCHILD, M. (Editors). **Crowdsourcing Geographic Knowledge**. Springer Dordrecht Heidelberg, New York, London, 2013. p. 245-263. ISBN 9789400745872

CONCAR. COMISSÃO NACIONAL DE CARTOGRAFIA. **Plano de Ação para Implantação da Infraestrutura Nacional de Dados Espaciais – INDE**, 2010. Disponível em: <http://www.concar.gov.br/pdf/PlanoDeAcaoINDE.pdf>. Acesso em: 11/12/2017.

CONCAR. COMISSÃO NACIONAL DE CARTOGRAFIA. **Perfil de Metadados Geoespaciais do Brasil (Perfil MGB)**, 2º edição, 2011. Disponível em: [http://www.concar.gov.br/pdf/111@Perfil\\_MGB\\_homologado\\_Nov\\_2009\\_v1.pdf](http://www.concar.gov.br/pdf/111@Perfil_MGB_homologado_Nov_2009_v1.pdf). Acesso em: 4/11/2017.

COPPOCK, J. T. & RHIND, D. W. The History of GIS. In: MAGUIRE, M. F.; GOODCHILD, M. F.; RHIND, D. W. **Geographical Information Systems: Principles and Applications**. Harlow, U.K.: Longman Group. 1991. p. 21-43.

CRUSE, D. A. The pragmatics of lexical specificity. **Journal of Linguistics**, v. 13, n. 2, September 1977. p. 153-164. Disponível em: <https://doi.org/10.1017/S0022226700005363>.

CURITIBA. PREFEITURA MUNICIPAL DE CURITIBA. Estrutura Ciclovária. Curitiba, 2017. Disponível em: <http://www.curitiba.pr.gov.br/conteudo/mais-bici-estrutura-ciclovitaria/2221>. Acesso em: 28/05/2017.

CURITIBA. Curitiba em Dados. Disponível em: <<http://curitibaemdados.ippuc.org.br>>. Acesso em: 12/02/2018.

D'ALGE, J. C. L. & GOODCHILD, M. F. Generalização cartográfica, representação do conhecimento e SIG. **Anais do VIII Simpósio Brasileiro de Sensoriamento Remoto**, Salvador, Bahia, 1996, p.147-151.

D'ALGE, J.C.L. **Generalização cartográfica em sistemas de informação geográfica: aplicação aos mapas de vegetação da Amazônia Brasileira**. Tese (Doutorado). Departamento de Engenharia de Transportes, Escola Politécnica da Universidade de São Paulo (USP), São Paulo, 2007. Disponível em: [http://www.nossoambiente.gov.br/estruturas/sbf\\_chm\\_rbbio/\\_arquivos/generalizacao\\_cartografica\\_em\\_sistemas\\_de\\_informacao\\_geografica\\_aplicacao\\_aos\\_mapas\\_de\\_vegetao\\_da\\_amaznia\\_brasileira.pdf](http://www.nossoambiente.gov.br/estruturas/sbf_chm_rbbio/_arquivos/generalizacao_cartografica_em_sistemas_de_informacao_geografica_aplicacao_aos_mapas_de_vegetao_da_amaznia_brasileira.pdf). Acesso em: 10/08/2017.

DAVIS JR., C. A. e LAENDER, A. H. F. Extensões ao Modelo OMT-G para Produção de Esquemas Dinâmicos e de Apresentação. In: Workshop Brasileiro de Geoinformática - GEOINFO 2000, 2000, São Paulo. **Anais GEOINFO 2000**, São Paulo, 2000. p. 29-36.

DNER. DEPARTAMENTO NACIONAL DE ESTRADAS DE RODAGEM. **Manual de Pavimentação (IPR-697)**. 2ª ed., Rio de Janeiro, 1996. 320p.

DNIT. DEPARTAMENTO NACIONAL DE INFRAESTRUTURA TERRESTRE. **Glossário de Termos Técnicos Rodoviários (IPR-700)**. 2ª ed., Rio de Janeiro, 2017. 324p.

DNIT. DEPARTAMENTO NACIONAL DE INFRAESTRUTURA TERRESTRE. **Manual de Pavimentação (IPR-719)**. 3ª ed, Rio de Janeiro, 2006. 274p.

DNIT. DEPARTAMENTO NACIONAL DE INFRAESTRUTURA TERRESTRE. **Manual de Implantação Básica de Rodovia (IPR-742)**. 3ª ed, Rio de Janeiro, 2010. 617p.

DNIT. DEPARTAMENTO NACIONAL DE INFRAESTRUTURA TERRESTRE. **Terminologias Rodoviárias Usualmente Utilizadas**. v. 1.1, 2007. 17p.

DNIT. DEPARTAMENTO NACIONAL DE INFRAESTRUTURA TERRESTRE. **Nomenclatura das rodovias federais**. Disponível em: <<http://www.dnit.gov.br/rodovias/rodovias-federais/nomeclatura-das-rodovias-federais>>. Acesso em: 13/01/2020.

DOAN, A., MADHAVAN, J.; DHAMANKAR, R.; DOMINGOS, P.; HALEVY, A. Learning to match ontologies on the Semantic Web. **The VLDB Journal - The International Journal on Very Large Data Bases**, v. 12, n. 4, p. 303-319, 2003.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **Manual Técnico T 34-700 Convenções Cartográficas (1ª Parte) Normas para o Emprego dos Símbolos**, 2ª Edição, 1998. Disponível em: [http://www.geoportal.eb.mil.br/images/stories/t34700\\_1aparte.pdf](http://www.geoportal.eb.mil.br/images/stories/t34700_1aparte.pdf). Acesso realizado em: 4/11/2017.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **Manual Técnico T 34-700 Convenções Cartográficas (2ª Parte) Catálogo de Símbolos**, 2ª Edição, 2000. Disponível em: [http://www.geoportal.eb.mil.br/images/PDF/T\\_34700\\_P2.pdf](http://www.geoportal.eb.mil.br/images/PDF/T_34700_P2.pdf). Acesso em: 4/11/2017.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-EDGV 2.1.3. – Especificação Técnica para a Estruturação dos Dados Geoespaciais Vetoriais**, 2010. Disponível em: [http://www.geoportal.eb.mil.br/portal/images/PDF/ET\\_EDGV\\_Vs\\_2\\_1\\_3.pdf](http://www.geoportal.eb.mil.br/portal/images/PDF/ET_EDGV_Vs_2_1_3.pdf). Acesso em: 30/03/2020.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-EDGV F-Ter. – Especificação Técnica para a Estruturação dos Dados Geoespaciais Vetoriais - 1ª parte**, 2016a. Disponível em: [http://www.geoportal.eb.mil.br/portal/images/PDF/EDGV\\_DefesaForca\\_Terrestre\\_2015.pdf](http://www.geoportal.eb.mil.br/portal/images/PDF/EDGV_DefesaForca_Terrestre_2015.pdf). Acesso realizado em: 30/03/2020.



DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-CQDG – Especificação Técnica para Controle de Qualidade de Dados Geoespaciais**, 2016b. Disponível em: <http://www.geoportal.eb.mil.br/index.php/inde2?id=142>. Acesso em: 3/10/2017.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-PCDG - Especificação Técnica para Produtos de Conjuntos de Dados Geoespaciais**, 2016c. Disponível em: [http://www.geoportal.eb.mil.br/portal/images/PDF/ET\\_PCDG\\_2016\\_2aEdicao\\_Aprovada\\_Publicada\\_BE\\_7\\_16.pdf](http://www.geoportal.eb.mil.br/portal/images/PDF/ET_PCDG_2016_2aEdicao_Aprovada_Publicada_BE_7_16.pdf). Acesso em: 27/04/2020.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-EDGV 3.0 – Especificação Técnica para a Estruturação dos Dados Geoespaciais Vetoriais**, 2017. Disponível em: <http://www.geoportal.eb.mil.br/index.php/inde2?id=142>. Acesso em: 02/11/2017.

DSG. DIRETORIA DO SERVIÇO GEOGRÁFICO – EXÉRCITO BRASILEIRO. **ET-ADGV – Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais 3.0**, 2018. Disponível em: <http://www.geoportal.eb.mil.br/index.php/inde2?id=142>. Acesso em: 12/01/2020.

ELIAS, E. N. N.; JESUS, E.G.V.; FERNANDES, V. O. Avaliação da Dispersão e Heterogeneidade de Dados Colaborativos do Openstreetmap. **Anais do VII Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação**, Recife - PE, 08-09 de Nov de 2018.

ELZAKKER, C. P. J. M. Van. **The Use of Maps in the Exploration of Geographic Data**. Tese de doutoramento, Netherlands Geographical Studies 326, ITC, Utrecht/Enschede, 2004.

ERIKSSON, H.; HARRIE, L.; PAASCH, J. M.; PERSSON, A. Techniques for and consequences of using INSPIRE extensions: a case study with Swedish hydrological data. **International Journal of Spatial Data Infrastructures Research**, v.13, p. 172-201, Special Section: INSPIRE (Full Research Article) 172, 2018.

ESTES, J. E. & MOONEYHAN, D. W. Of Maps and Myths. **Photogrammetric Engineering and Remote Sensing**. v. 60, p. 517-524, 1994.

ESTIMA, J. E PAINHO, M. Exploratory analysis of OpenStreetMap for land use classification. **Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information, GEOCROWD '13**. November, 2013. p. 39-46.

FAHAD, M. ER2OWL: Generating OWL ontology from ER diagram. **Intelligent Information Processing IV, 5th IFIP International Conference on Intelligent Information Processing**, October 19-22, 2008, Beijing, China.

FAIRBAIN, D. J. The frontier of cartography: mapping a changing discipline. **Photogrammetric Record**, v.14, n. 84, p. 903-915,1994.

FALBO, R. A. **Integração de Conhecimento em um Ambiente de Engenharia de Software**. Tese (doutorado), 215 p. Departamento de Engenharia de Sistemas e Computação, Universidade Federal do Rio de Janeiro, Rio de Janeiro-RJ, 1998.

FELICÍCIMO, Carolina Howard. **Interoperabilidade semântica na web: uma estratégia para o alinhamento taxonômico de ontologias**. Dissertação (mestrado), 180 p. Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, Rio de Janeiro-RJ, 2004.

FLANAGIN, A. J.; METZGER, M. J. The credibility of volunteered geographic information. **GeoJournal**, v. 72, p. 137-148, 2008, Disponível em: <https://pdfs.semanticscholar.org/78ed/f53204beec5b77b2f895719d354cf1f5065f.pdf>. Acesso em: 6/01/2018.

FONSECA, F., EGENHOFER, M., BORGES, K. A. Ontologias e Interoperabilidade Semântica entre SIGs. In: **Proceedings do II Workshop Brasileiro em Geoinformática (GeoInfo2000)**, 2000. São Paulo.

FOODY, G.M; SEE L.; FRITZ, S.; VAN DER VELDE M.; PERGER, C.; SCHILL C.; BOYD, D. S.; COMBER, A. Accurate attribute mapping from volunteered geographic information: issues of volunteer quantity and quality. **Cartographic Journal**, v. 52, 2015.

GAEVI, D.; DJURI, D.; DEVEDI, V. & DAMJA, V. Converting UML to OWL ontologies. **Proceedings of the 13th international World Wide Web Conference**, 2004.

GIRRES, J-F & TOUYA, G. Quality assessment of the French OpenStreetMap dataset. **Transactions in GIS**, v. 14, p.435-459, 2010.

GOLLEDGE, R.; STIMSON, R. J. **Analytical Behavioural Geography**. New York: Croom Helm, 1987.

GOODCHILD, M. F. **Geographical data modeling**. **Computers & Geosciences**, v.18, n. 4, p. 401-408, 1992.

GOODCHILD, M. F. Citizens as sensors: the world of volunteered geography. **GeoJournal**, v. 69, p. 211-221, 2007.

GOODCHILD, M. F.; LI, L. Assuring the quality of volunteered geographic information. **Spatial Statistics**, v.1, p.110-120, 2012. Disponível em: <https://helios.geog.ucsb.edu/~good/275/vgiquality.pdf>. Acesso em: 6/01/2018.

GOODCHILD & LONGLEY. The future of GIS and spatial analysis. In: **Geographical Information Systems: Principles, Techniques, Management and Applications**, 2nd Edition, 404 p, Abridged. (Eds.) LONGLEY, P.A.; GOODCHILD, M. F.; David J. MAGUIRE, D. J.; RHIND, R. W., 2005. ISBN: 978-0-471-73545-8.

GOODCHILD, M. NeoGeography and the nature of geographic expertise. **Journal of Location Based Services**, v. 3, n. 2, p. 82-96, 2009.

GRÖCHENIG, S., BRUNAUER, R. & REHRL, K. Digging into the history of VGI datasets: results from a worldwide study on OpenStreetMap mapping activity. **Journal of Location Based Services**, v. 8, n. 3, p.198-210, 2014.

GRUBER, T. R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. **International Journal Human-Computer Studies**, v. 43, p.907-928,1993a.

GRUBER, R. T. A **Translation Approach to Portable Ontology Specifications**. Knowledge Acquisition, v. 5, n. 2, p. 199-220, 1993b.

GUARINO, N. (ed.) Formal Ontology in Information Systems. **Proceedings of FOIS'98**, Trento, Italy, Amsterdam IOS Press, 6-8 June 1998. p. 3-15.

GITHUB. Disponível em: <https://github.com>. Acesso em: 20 de Janeiro de 2020.

GUIZZARDI, G. **Ontological foundations for structural conceptual models**. PhD Thesis Series, 416 p. Enschede: Telematica Instituut (CTIT), 2005.

GUPTILL, S. C. Metadata and data catalogues. **In: Geographical Information Systems: Principles, Techniques, Management and Applications**, 2nd Edition, Abridged. (Eds) LONGLEY, P. A.; GOODCHILD, M.F.; MAGUIRE, D. J., David W. RHIND, D. W. May, 2005, 404 pgs. ISBN: 978-0-471-73545-8,

HAENDCHEN, F. A.; STAA, A. V.; LUCENA, C. P. A. Component-Based Model for Building Reliable Multi-Agent Systems. **Anais do SEW - NASA/IEEE Software Engineering Workshop**. Los Alamitos, 2003.

HAGEDORN, T. J., SMITH, B., KRISHNAMURTY, S. e GROSSE, I. Interoperability of disparate engineering domain ontologies using basic formal ontology. **Journal of Engineering Design**, 2019.

HAKLAY, M. & WEBER, P. OpenStreetMap:User-Generated Street Maps. **IEEE Pervasive Computing**, v. 7, p. 12-18, 2008.

HAKLAY, M. How Good is volunteered geographical information? a comparative study of OpenStreetMap and ordnance survey datasets. **Environment and Planning B: Planning and Design**, v.37, n. 4, p. 682-703, 2010.

HAMPTON, J. A. Polymorphous Concepts in Semantic Memory. **Journal of Verbal Learning and Verbal Behavior**, v.18, n. 4, p.441-461, 1979.

HAMPTON, J. A. An investigation of the nature of abstract concepts. **Memory & Cognition**, v.9, n.2, p. 149-156, 1981.

HAYES-ROTH, F.; WATERMAN, D.; LENAT, D.. **Building Expert Systems**. Addison-Wesley, 472 p., 1983. ISBN 0-201-10686-8.

HEIPKE, C. Crowdsourcing geospatial data. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 65, p. 550–557, 2010.

HOVY, E.; NAVIGLI, R.; PONZETTO, S. P. Collaboratively built semi-structured content and Artificial Intelligence: The story so far. **Artificial Intelligence**, v. 194, p. 2-27, 2013.

HU, Y.; MCKENZIE, G.; JANOWICZ, K. & GAO, S. Mining Human-Place Interaction Patterns from Location-Based Social Networks to Enrich Place Categorization Systems. **Proceedings of the Workshop on Cognitive Engineering for Spatial Information Processes at COSIT**. Santa Fe, 2015.

HU, Y. Geospatial Semantics. In: (Eds) HUANG, B.; COVA, T. J. & TSOU, M. *et al.* **Comprehensive Geographic Information Systems**, Elsevier. Oxford, UK, 2017. Disponível em: <http://dx.doi.org/10.1016/B978-0-12-409548-9.09597-X>. Acesso em: 29/11/2019.

IBGE. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Classificação Nacional de Atividades Econômicas (CNAE) Subclasses v. 2.2**. Janeiro, 2015.

IBGE. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Guia das atividades de Geociências do IBGE para Jornalistas**. 51p. Ed. IBGE, Diretoria de Geociências, Rio de Janeiro, 2015. Disponível em: <https://biblioteca.ibge.gov.br/biblioteca-catalogo.html?id=293539&view=detalhes>. Acesso em: 16/12/2017d. ISBN: 2409353700.

IBGE. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. COORDENAÇÃO DE CARTOGRAFIA. **Avaliação da qualidade de dados geoespaciais** (Manuais técnicos em geociências, ISSN 0103-9598; n. 13) Rio de Janeiro, 2017, 96 p. ISBN 978-85-240-4423-6.

IBGE. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. BASES CARTOGRÁFICAS CONTÍNUAS – BRASIL. **Base Contínua do Estado do Rio de Janeiro na Escala 1:25.000**, 2018. Disponível em: <https://www.ibge.gov.br/geociencias/cartasemapas/basescartograficascontinuas/15759-brasil.html?=&t=downloads>. Acesso em: 09/02/2020.

INDE. INFRAESTRUTURA NACIONAL DE DADOS ESPACIAIS. **Plano de Ação para Implantação da Infraestrutura Nacional de Dados Espaciais**. Rio de Janeiro, 2010. Disponível em: <http://www.concar.gov.br/pdf/PlanoDeAcaoINDE.pdf>. Acesso em: 23/10/2017.

INDE. INFRAESTRUTURA NACIONAL DE DADOS ESPACIAIS. Disponível em: [http://www.inde.gov.br/images/inde/Folder\\_INDE\\_port\\_jul2011.pdf](http://www.inde.gov.br/images/inde/Folder_INDE_port_jul2011.pdf). Acesso em: 23/10/2017.

IPPUC. INSTITUTO DE PESQUISA E PLANEJAMENTO URBANO DE CURITIBA. **Dados do Sistema Viário Classificado**, 2015. Disponível em: <http://ippuc.org.br/geodownloads/geo.htm>. Acesso em: 10/02/2020.

ISO/TC211. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 19113:2002 Geographic information – Quality principles**, 2002.

ISO/TC211. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 19114:2003 Geographic information – Quality evaluation procedures**, 2003.

ISO/TC211. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 19115:2003 Geographic Information – Metadata**, 2015.

ISO/TC211. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 19125-1:2004 Geographic information — Simple feature access — Part 1: Common architecture**, 2004.

ISO/TC211. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 19125-2:2004 Geographic information — Simple feature access — Part 2: SQL option**, 2004.

ISO/TC211. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 19138:2006 Geographic information – Data quality measures**, 2006.

ISO/TC211. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 19157:2013 Geographic Information – Data Quality**, 2013.

JANOWICZ K.; SCHEIDER S.; ADAMS B. **A Geo-semantics Flyby**. In: (Eds) RUDOLPH S.; GOTTLOB G.; HORROCKS, I.; VAN HARMELEN; F. **Reasoning Web: Semantic Technologies for Intelligent Data Access. Reasoning Web 2013. Lecture Notes in Computer Science**, v. 8067. Springer, Berlin, Heidelberg, p. 230-250, 2013.

JARKE, M., NEUMANN, B., VASSILIOU, Y. & WAHLSTER, W. KBMS Requirements of Knowledge-Based Systems. In: (Eds.) SCHMIDT, J. W. & THANOS, C.: **Foundations of Knowledge Base Management: Contributions from Logic, Databases, and Artificial Intelligence**. Berlin: Springer, 1989. p. 391-195.

JOKSIĆ D.; BAJAT B. Elements of spatial data quality as information technology support for sustainable development planning. **Spatium**, n. 11, Institute of Architecture and Urban & Spatial Planning of Serbia, Belgrade, p.77-83, 2004.



JONES, D.; BENCH-CAPON, T.; VISSER, P. Methodologies for ontology development. In: **Proceedings of IT KNOWS Conference**, XV IFIPWorld Computer Congress, Budapest, August, 1998.

KEATES, J. S. **Projecto e produção cartográfica**. 240 p., 1973.

KIFER, M.; LAUSEN, G. & WU, J. Logical foundations of object-oriented and framebased languages. **Journal of the ACM**, n. 42, p. 741-843, 1995.

KSL - **Knowledge Systems Laboratory – Stanford University**. Disponível em: <http://www-ksl-svc.stanford.edu/>. Acesso em: 13/12/2019.

KUHN, W. Geospatial Semantics: Why, of What, and How? **Journal on Data Semantics III**, LNCS, n. 3534, p. 1-24, 2005.

LACASTA, J.; LOPEZ-PELLICER, F. J.; RENTERIA-AGUALIMPIA, W. & NOGUERAS-ISO, J. Improving the visibility of geospatial data on the Web. **IEEE/ACM Joint Conference on Digital Libraries**, 2014.

LAKOFF, G. **Woman, fire, and dangerous things: what categories reveal about the mind**. Chicago: University of Chicago Press, 1987.

LIMA, P.; CÂMARA, G.; QUEIROZ, G. **GeoBR: Intercâmbio Sintático e Semântico de Dados Espaciais**. INPE, São José dos Campos, SP, Brasil, 2002.

LLOYD, R.; PATTON, D; CAMMACK, R. Basic-level Geographic Categories. **The Professional Geographer**, v. 48, n. 2, 1996.

LOBBEN, A. K. Tasks, strategies, and cognitive processes associated with navigational map reading: a review perspective. **The Professional Geographer**, v. 56, n. 2, p. 270-281, 2004.

LUDWIG, C. & ZIPF, A. Exploring regional differences in the representation of urban green spaces in OpenStreetMap. **Proceedings of the “Geographical and Cultural**

**Aspects of Geo-Information: Issues and Solutions” AGILE 2019 Workshop**, Limassol, Cyprus, June 17th 2019.

MACEACHREN, A. M. The role of maps in spatial knowledge acquisition. **The Cartographic Journal**, London, v. 28, p. 152-162, 1991.

MACEACHREN, A. M. **How maps work: representation, visualization, and design**. New York: The Guilford Press, 1995.

MALONE, T. W. What is collective intelligence and what will we do about it? In: (Ed.) MARK, T. **Collective intelligence: creating a prosperous worlds at peace**. Oaktown: Earth Intelligence Network, p. 37-40, 2008.

MALONE, T. W.; LAUBACHER, R.; DELLAROCAS, C. Harnessing Crowds: Mapping the Genome of Collective Intelligence. In: **MIT Sloan Research Paper**, v. 4732-09, 2009.

MARK, D.; EGENHOFER, M.; HIRTLE, S. & SMITH, B. **UCGIS emerging research theme: Ontological foundations for geographic information science**. In: A Research Agenda for Geographic Information Science, 2000. p.335-350.

MASÓ, J.; TORRES, A. Z.; BAUMANN, P. New Model for Geospatial Coverages in JSON: Coverage Implementation Schema and Its Implementation With JavaScript. **Emerging Technologies and Applications in Data Processing and Management**, p. 316-357, 2019.

MERVIS, C. B. & ROSCH. E. Categorization of natural objects. In: **Annual review of psychology**, (Eds.) ROSENSWEIG, M. R. and PORTER, L. W., Palo Alto, CA: Annual Reviews, Inc. 32, 1981. p. 89-115.

MICHAELIS DICIONÁRIO ONLINE. Disponível em: <<https://michaelis.uol.com.br/modernoportugues/busca/portuguesbrasileiro/sem%C3%A2ntica/>>. Acesso em: 30/12/2019.

MKHININI, M.M.; LABBANI-NARSIS, O.; NICOLLE, C. Combining UML and ontology: An exploratory survey. **Computer Science Review**, v. 35, 2020.

MONTELLO, D. Cognitive Map-Design Research in the Twentieth Century: Theoretical and Empirical Approaches. **Cartography and Geographic Information Science**, v. 29, n. 3, 2002.

MORAIS, E. A. M. & AMBRÓSIO, A. P. L. **Ontologias: conceitos, usos, tipos, metodologias, ferramentas e linguagens**. Technical Report - INF\_001/07 -, Dezembro, 2007. Relatório Técnico.

MURPHY, G. L. & SMITH, E. E. Basic-Level Superiority in Picture Categorization. **Journal of Verbal Learning and Verbal Behavior**, v. 21, p. 1-20, 1982.

NOVACK, T.; VOSS, J.; SCHULTZ, M.; ZIPF, A. Associating OpenStreetMap tags to CORINE land-cover classes using text and semantic similarity measures. In: **VGI-ALIVE Workshop at AGILE**, Helsinki, 2016.

NOVACK, T.; PETERS, R.; ZIPF, A. Graph-Based Matching of Points-of-Interest from Collaborative Geo-Datasets. **ISPRS International Journal Geo-Information**. v.7 , n.117, p. 1-17, 2018.

NOVACK, T.; GRINBERGER, A. Y.; SCHULTZ, M.; ZIPF, A.; MOONEY, P. The Geographical and Cultural Aspects of Geo-Information: An Introduction. **Proceedings of the “Geographical and Cultural Aspects of Geo-Information: Issues and Solutions” AGILE 2019 Workshop**, Limassol, Cyprus, June 17th, 2019.

NOY, N. F. & MUSEN, M. A. SMART: Automated Support for Ontology Merging and Alignment. In: **Workshop on Knowledge Acquisition, Modeling and Management**. Banff, 1999.

NOY, N. F. & MUSEN, M. A. Anchor-PROMPT: Using Non-Local Context for Semantic Matching. **Workshop on Ontologies and Information Sharing at the**

**Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-2001)**, Seattle-CA, 2001.

NOY, N. F. & MUSEN, M. A. The PROMPT Suite: Interactive Tools For Ontology Merging And Mapping. **International Journal of Human-Computer Studies**, v. 59, n. 6, p.983-1024, 2003.

NUSEIBEH, B.; EASTERBROOK, S.; RUSSO, A. Leveraging Inconsistency in Software Development. **IEEE Computer Society Press**, v. 33, n. 4, p. 26-29, 2000.

OGC. OPEN GEOSPATIAL CONSORTIUM. Disponível em:<<http://www.opengeospatial.org/>>. Acesso em: 24/01/2020.

OGC. OPEN GEOSPATIAL CONSORTIUM. **Geography Markup Language (GML)**. Disponível em: <https://www.opengeospatial.org/standards/gml>. Acesso em: 14/12/2019.

OLTEANU-RAIMOND, A.-M.; HART, G.; FOODY, G.; TOUYA, G.; KELLENBERGER, T.; DEMETRIOU, D. The scale of VGI in map production: A perspective of European National Mapping Agencies. **Transactions in GIS**, v. 21, p. 74-90, 2017.

OSM. OPEN STREET MAP. Disponível em: <https://www.openstreetmap.org/about>. Acesso em: 27/01/2020.

OSM FOUNDATION. OPEN STREET MAP FOUNDATION. Disponível em: <http://wiki.osmfoundation.org/wiki/>. Acesso em: 27/01/2020.

PEREIRA, K. D.; AUGUSTO, M. J. C.; SANTOS, C. J. B.; FREITAS, A. L. Atualização da Legislação Cartográfica: Necessidade Nacional. **Anais do XXI Congresso Brasileiro de Cartografia**. Belo Horizonte, 2003. Disponível em: <[www.ibge.gov.br/.../2003/XXI\\_Congresso\\_Brasileiro\\_de\\_Cartografia/CBC\\_2003\\_WORK\\_GT6\\_D.pdf](http://www.ibge.gov.br/.../2003/XXI_Congresso_Brasileiro_de_Cartografia/CBC_2003_WORK_GT6_D.pdf)>. Acesso em: 17/12/2017.

PETERSON, M. P. **Interactive and animated cartography**. Prentice Hall, New Jersey, 257p., 1995.

PINHO, C. M. D. & GOLTZ, E. **Construção de ontologias espaciais: O lote urbano**. Trabalho (Disciplina Banco de Dados Geográficos) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2003. Disponível em: [www.dpi.inpe.br/livros/bdados/artigos/lote.pdf](http://www.dpi.inpe.br/livros/bdados/artigos/lote.pdf). Acesso em: 12/12/2019.

PINTO, H. S.; GÓMEZ-PÉREZ, A.; MARTINS, J. P. Some Issues on Ontology Integration. **Proceedings of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods (KRR5)**, (Eds.) BENJAMINS, V.R.; CHANDRASEKARAN, B.; GÓMEZ-PÉREZ, A.; GUARINO, N.; USCHOLD, M. Stockholm, Sweden, 1999.

POTONIEC, J.; WISNIEWSKI, D.; LAWRYNOWICZ, A.; KEET, C. M. Dataset of ontology competency questions to SPARQL-OWL queries translations. **Data in Brief**, v. 29, 2020.

PROTÉGÉ. Disponível em: <https://protege.stanford.edu/>. Acesso em: 13/12/2019.

RAK, A. **Legal Issues and Validation of Volunteered Geographic Information**. M.Sc.E. Thesis, 128 p. Department of Geodesy and Geomatics Engineering Technical Report N°. 283, University of New Brunswick, Fredericton, New Brunswick, Canada, 2013.

RAYMOND, E. S. **The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary**. O'Reilly, Beijing, 1999.

RODRÍGUEZ DÍAZ, A.; BENITO-SANTOS, A.; DORN, A.; ABGAZ, Y.; WANDL-VOGT, E.; THERÓN, R. Intuitive Ontology-Based SPARQL Queries for RDF Data Exploration. **IEEE Access**, v. 7, 2019.

RÖNNEBERG, M., LAAKSO, M., SARJAKOSKI, T. Map Gretel: social map service supporting a national mapping agency in data collection. **Journal of Geographical Systems**, v. 21, p. 43-59, 2019.

ROSCH, E.; MERVIS, C.B.; GRAY, W.D.; JOHNSON, D.M.; BOYES-BRAEM, P. Basic Objects in Natural Categories. **Cognitive Psychology**, n. 8, p. 382-439, 1976.

ROSCH, E. Natural categories. **Cognitive Psychology**, n. 4, 1973.

ROSCH, E. Principles of Categorization. In: (Eds.) ROSCH, E.; LLOYD, B.B. **Cognition and Categorization**. Hillsdale: Erlbaum, 1978.

ROSH, E. & MERVIS, C. B. Family resemblances: Studies in the internal structure of categories. **Cognitive Psychology**, v. 7, n. 4, pp. 573-605, October, 1975.

SANTIL, F. L. P. & SLUTER, C. R. As pesquisas em cognição visual aplicadas à Cartografia. **Revista Brasileira de Cartografia**, v. 64, n. 3, p. 367-376, 2012.

SIEBER, R. E. & JOHNSON, P. A. Situating the Adoption of VGI by Government. In: (Eds.) SUI, D. Z.; ELWOOD, S.; GOODCHILD, M. **Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice**. Springer Science+Business Media, Dordrecht, 2013.

SILVA, V. A.; MACHADO, A. A.; CAMBOIM, S. P. **Ferramenta para Importação de Dados do OpenStreetMap para um Banco de Dados de Referência Estruturado de Acordo com o Padrão ET-EDGV**. No prelo.

SLUTER, C. R.; VAN ELZAKKER C. P. J. M.; IVÁNOVÁ, I. Requirements Elicitation for Geo-Information Solutions. **The Cartographic Journal**, The British Cartographic Society, p. 1-14, 2016.

SLUTER, C. R.; BRANDALIZE, M.C.B.; IVÁNOVÁ, I.; VAN ELZAKKER, C. P. J. M. Projeto Cartográfico Centrado no Usuário dos Mapas Necessários à Definição da Estrutura do Sistema Viário de Um Plano Diretor Municipal. **Revista Brasileira de Cartografia**, v. 67, n. 8, p. 1557-1567, 2015.

SLUTER, C. R. Território. **In: Brasil em números**. IBGE, Rio de Janeiro, v. 21, p. 43-52, 2013.

SMITH, B. e MARK, M. Geographical categories: an ontological investigation. **International Journal of Geographical Information Science**, v. 15, n. 7, p. 591-612, 2001.

SMITH, B. e MARK, D. M. Ontology and geographic kinds. **Proceedings, International Symposium on Spatial Data Handling**, Vancouver, Canada. 12-15 July, 1998.

SOARES, D. & D'ALGE, J. A Escala na Representação de Elementos em um Banco de Dados Geográfico: Conceitos, Implicações e Uso. **IV Colóquio Brasileiro de Ciências Geodésicas - IV CBCG**. Curitiba-PR, 16 a 20 de maio de 2005.

SONDHEIM, M.; GARDELS, K. & BUEHLER, K. GIS interoperability. **In: Geographical Information Systems: Principles, Techniques, Management and Applications**. (Eds.) LONGLEY, P.A.; GOODCHILD, M.F.; Maguire, D.J.; RHIND, D.W. 2nd Edition, 404p., Abridged., May 2015.

SOUSA, J. L.; MARTINS, P. G. M.; RAMALHO, R. A. S. Modelos de representação semântica na era do Big Data. **Brazilian Journal of Information Studies: Research Trends**, v. 12, n. 3, p.34-40, 2018.

STAAB, S; SCHNURR, H; STUDER, R; SURE, Y. Knowledge processes and ontologies. **IEEE Intelligent Systems**, v.16, n.1, p. 26-34, 2001.

STEPHEN, S.; & HAHMANN, T. Formal Qualitative Spatial Augmentation of the Simple Feature Access Model. **In: 14th International Conference on Spatial Information Theory (COSIT 2019)**. (Eds) TIMPF, S.; SCHLIEDER, C.; KATTENBECK, M.; LUDWIG, B. and STEWART, K. Article n. 15 p. 15:1–15:18. Leibniz International Proceedings in Informatics Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany.



SURE Y., ANGELE J., STAAB, S. OntoEdit: Multifaceted Inferencing for Ontology Engineering. In: (Eds) SPACCAPIETRA, S.; MARCH, S.; ABERER, K. **Journal on Data Semantics I**. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg. v. 2800, 2003.

SUROWIECKI, J. **The wisdom of crowds**. First Anchor Books Edition. New York: Anchor Books, 2005.

SUROWIECKI, J. **The collective intelligence of the Web**. The New Yorker, May 18, 2014.

THORNDYKE, P. W. & HAYES-ROTH, B. Differences in spatial knowledge acquired from maps and navigation. **Cognitive Psychology**, v.14, p. 560-589, 1982.

TOMLINSON, R. F. & PETCHENIK, B. B. (Eds.) Reflections on the revolution: The transition from analogue to digital representations of space. **The American Cartographer**, IS (3) special issue, p. 1958-1988, 1988.

TOUYA, G.; ANTONIOU, V.; CHRISTOPHE, S.; SKOPELITI, A. Production of Topographic Maps with VGI: Quality Management and Automation. In: (Eds.) FOODY, G.; SEE, L.; FRITZ, S.; MOONEY, P.; OLTEANU-RAIMOND, A-M.; FONTE, C.C. & ANTONIOU, V. **Mapping and the Citizen Sensor**. Ubiquity Press Ltd., London, p. 61-92, 2017.

TVERSKY, B. & HEMENWAY, K. Objects, Parts, and Categories. **Journal of Experimental Psychology: General**, v. 113, n. 2, 1984.

UFPR. Universidade Federal do Paraná. Base de Dados Científicos da Universidade Federal do Paraná. Disponível em: <https://bdc.c3sl.ufpr.br/>. Acesso em: 02/04/2020.

UREÑA-CÁMARA, M. A; NOGUERAS-ISO, J.; LACASTA, J. & ARIZA-LÓPEZ, F. J. A method for checking the quality of geographic metadata based on ISO 19157. **International Journal of Geographical Information Science**, 2018.

VAN NGUYEN, T., VO, H. L. M., HOANG, Q. e HOANG, H. H. A new method for transforming TimeER model based specification into OWL ontology. **8th Asian Conference on Intelligent Information and Database Systems ACIIDS**, Da Nang, Vietnam, 2016.

VEREGIN, H. Data quality parameters. In: LONGLEY, P.A.; GOODCHILD, M.F.; MAGUIRE, D.J. and RHIND, D.W. (Eds), **Geographical information systems**. New York, Wiley, p. 177-189, , 1999.

VO, M. H. L. & HOANG, Q. Transformation of UML class diagram into OWL Ontology. **Journal of Information and Telecommunication**, v. 4, n.1, p. 1-16, 2019.

W3C. WORLD WIDE WEB CONSORTIUM. **XML Essentials**. Disponível em: <<https://www.w3.org/standards/xml/core>>. Acesso realizado em 26/01/2020.

WEBER, E.; ANZOLCH, R.; JUGURTA, L.F.; COSTA, A. C.; IOCHPE, C. **Qualidade de Dados Geoespaciais**. (RP- n°293-99), Porto Alegre: Instituto de Informática, UFRGS, 1999.

WILKINSON, D. M. & HUBERMAN, B. A. Assessing the Value of Cooperation in Wikipedia. **First Monday**, v. 12, n. 4, 2007. Disponível em: <<http://arxiv.org/abs/cs/0702140>>. Acesso em: 01/01/2018.

WIKI OSM. WIKI OPEN STREET MAP. **API v 0.6**. Disponível em: [https://wiki.openstreetmap.org/wiki/API\\_v0.6](https://wiki.openstreetmap.org/wiki/API_v0.6). Acesso em: 27/01/2020.

WIKI OSM. WIKI OPEN STREET MAP. **Press Kit**. Disponível em: [http://wiki.openstreetmap.org/wiki/Pt:Press\\_Kit](http://wiki.openstreetmap.org/wiki/Pt:Press_Kit). Acesso em: 27/01/2020b.

WIKI OSM. WIKI OPEN STREET MAP. **Elements**. Disponível em: <http://wiki.openstreetmap.org/wiki/Elements>. Acesso em: 8/11/2017.